Ontology driven framework for multimedia information retrieval in P2P network

Maria Sokhn

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Informatique et Réseaux »

présentée et soutenue publiquement par

Maria SOKHN

26 Août 2011

Plateforme de recherche basée sur l'ontologie pour des données multimédia dans un réseau pair à pair

Directeur de thèse : **Ahmed SERHROUCHNI**
Co-encadrement de la thèse : **Omar ABOU KHALED**

Jury

Mme Elena Mugellini, Pr., iTIC, Ecole díngénieurs et d'architecte de Fribourg	Présidente
M. Philippe CUDRE-MAUROUX, Pr., Diuf, Université de Fribourg	Rapporteur
Mme Rima KILANY, Maître de conférence, ESIB, Université Saint Joseph	Rapporteur
M. Jean-Marc OGIER, Pr., L3i de recherche, Université La Rochelle	Examinateur
M. Samer LAHOUD, Maître de conférence, IRISA, IUT St. Malot	Examinateur
M. Jacques BERSIER, Responsable Ra&D, Ecole d'ingénieurs et d'architecte de Fribourg	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

THÈSE

To my parents,

Acknowledgements

First paper, first conference, first project, first course... A valuable experience which was made possible thanks to the wonderful people I met and to whom I would like to express my sincere acknowledgment.

I wish to thank first and foremost, Professor Ahmed Serhrouchni, my thesis advisor, who offered me the opportunity to enroll in Telecom ParisTech for my PhD thesis. I thank him for the trust he had in me and his unconditional help. I would like to thank Jean-Yves Le Meur, section leader at the CERN, who has made possible my enrollment through the financial support granted by the CERN.

Words cannot express my endless and sincere thanks to Professor Abou Khaled, my co-advisor. I thank him for having consistently encouraging me over the four years. I thank him for believing in me and pushing me to the best of my abilities. I thank him for his support, guidance, encouragement and advice. I was fortunate to have, also as a co-advisor, Professor Elena Mugellini from whom I received precious help all long my work. I thank her for all her constant availability, her constructive remarks and her involvement and motivation in pushing forward the activities around the PhD thesis work.

I wish to express my sincere thanks to the members of the jury. Specially Professor Rima Kilany and Professor Philippe Cudre-Mauroux for their precious time and valuable suggestions which helped improve my thesis report. Professor Jean-Marc Ogier for his commitment and availability, Professor Samer Lahoud for his encouragement, sense of humor and important advice, and Jacques Bersier who kindly agreed to be a member of my PhD thesis jury.

I address my special thanks to Michel Rast, director of the University of Applied Sciences of Western Switzerland, and Antoine Delley, director of the Information Technology and Communication Institute, who warmly welcomed me in their establishment where I spent four wonderful and memorable years of work and rich exchanges. I wish to extend my cordial appreciation to the engineers with whom I have closely worked with. Mainly Jean Revertera, David Godel, Dominic Wyler, Joel Dumoulin, Jonathan Stoppani and Samir Atittallah. I thank them for their time and valuable contribution.

I am indebted to my colleagues and friends who never stopped

supporting me, particularly Didier Perroud and Stuart Miller. *Didier, I am deeply grateful for your care, your attention, your friendship and all the "philosophical" discussions we had. Stuart, you may not know this but you helped me enormously and when I needed it the most, thank you.*

I owe my warmest thanks to Marck who made available his support in a number of ways. *Marck, I am infinitely thankful for your tireless encouragement, your calming presence, your incomparable support, your tenderness, your attention, your listening, your advice, your infinite patience. You always surprised me, calmed me down, made me laugh, took care of me. You knew how to deal with my caprices and black goat temperament. Merrrci!*

My loving thanks go to my wonderful parents, Michel and Ninette without whom I would not have dare thinking of achieving any of my dreams. *I am proud to be your daughter and blessed to have a wonderful brother and a delightful sister. Joe, Nayla, I love you !*

...I am writing this acknowledgment today, because, almost 5 years ago, at the Palmier hotel in Lausanne, I was encouraged by a wonderful person to jump into the adventure of a PhD thesis. *For having,* ***...all the way****, filled my heart with courage and confidence, I address you Jimmi, my deepest and most sincere gratitude.*

Publications

EDITED BOOK & PROCEEDINGS:

[1] *Advances in Intelligent Web Mastering - 3: Proceedings of the 7th Atlantic Web Intelligence Conference, AWIC 2011, Fribourg, Switzerland, January, 2011.* Elena Mugellini, Piotr S. Szczepaniak, Maria Chiara Pettenati, Maria Sokhn. Advances in Soft Computing series, Springer, 250 pages.

[2] *Special issue for the journal of Ambient Intelligence and Humanized Computing,* ***To be published on January 2012****.* Editor-in-Chief: Vincenzo Loia, Guest Editors: Elena Mugellini, Maria Sokhn, Piotr S. Szczepaniak. Journal no. 12652, Computational Intelligence and complexity, Springer, 70 pages.

JOURNAL:

[1] Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni. *Conference knowledge modeling for conference-video-recordings querying & visualization.* Journal of Multimedia Processing and Technologies Volume 1, Number 2, June 2010.

[2] [*in submission*] Francesco Carrino, Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Memoria-Mea: Personal Information Management.* Journal of Ambient Intelligence and Humanized Computing. May, 2011.

INVITED PAPER:

[1] Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni. *End-to-end adaptive framework for multimedia information retrieval*, 9th International Conference on Wired/Wireless Internet Communications (WWIC 2011), June, 2011, Vilanova i la Geltru, Catalonia, Spain.

PAPER:

[1] Samir Atitalah, Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Semantic multimedia management in P2P Network*, SEKE 2011, 23rd International Conference on Software Engineering and Knowledge Engineering, July, 2011, Miami Beach, USA.

[2] Joel Dumoulin, Maria Sokhn, Elena Mugellini, Omar Abou Khaled and Ahmed Serhrouchni, *Vers une approche de visualisation multivues dans un systeme d'information repartie*, NOTERE 2011, 11th Annual International Conference on New Technologies of Distributed Systems, May, 2011, Paris, France.

[3] Samir Atitallah, Tobias Wunden, Maria Sokhn, Elena Mugellini,

Omar Abou Khaled. *Speech to text in replay.* The 7th Atlantic Web Intelligence Conference. Advances in Soft Computing series, Springer. January, 2011, Fribourg, Switzerland.
[4] Jean Revertera, Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *SMAC: Smart Multimedia Archiving conference.* The 7th Atlantic Web Intelligence Conference. Advances in Soft Computing series, Springer. January, 2011, Fribourg, Switzerland.
[5] Francesco Carrino, Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *MEMORIA-MEA: Combining semantic technologies and interactive visualization techniques for personal information management.* The 7th Atlantic Web Intelligence Conference. Advances in Soft Computing series, Springer. January, 2011, Fribourg, Switzerland.
[6] Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Querying over heterogeneous and distributed data sources.* The 7th Atlantic Web Intelligence Conference. Advances in Soft Computing series, Springer. January, 2011, Fribourg, Switzerland.
[7] Maria Sokhn, Francesco Carrino, Elena Mugellini, Omar Abou Khaled. *Conference knowledge modeling for conference-video-recordings, querying and visualization.* ACM-MEDES, October, 2009, LYON, France.
[8] Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Knowledge modeling for enhanced information retrieval and visualization.* 6th Atlantic Web Intelligence Conference. Advances in Soft Computing series, Springer. September, 2009, Prague, Czech Republic.
[9] Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Knowledge management framework for conference video-recording retrieval.* The 21st International Conference on Software Engineering and Knowledge Engineering. July, 2009, Boston, USA.

POSTER:
[1] Joel Dumoulin, Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Multimedia information browsing and visualization.* IEEE VisWeek-InfoViz, October 2010, Salt Lake City, Utah, USA.
[2] Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni. *Conference advanced level information management and retrieval.* Spring School on Social Media Retrieval, February, 2010, Interlaken, Switzerland.
[3] Maria Sokhn, Elena Mugellini, Omar Abou Khaled. *Knowledge based video-recordings visualization.* The 8th International Semantic Web Conference, October, 2009, Washington DC, USA.

Administrative context

My Thesis was accomplished with the national school of telecommunication, based in Paris, FRANCE (Telecom ParisTech), in collaboration with the intergovernmental organization, based in Geneva, SWITZERLAND (CERN), and the university of applied sciences of western switzerland, based in Fribourg, SWITZERLAND (EIAFR).

SUBJECT TITLE: *Ontology driven framework for multimedia information retrieval in P2P networks.*

TOPICS: Multimedia Information retrieval, Knowledge modeling and management, Ontology, Semantic web, Model based retrieval and visualization, Retrieval in P2P network.

SPECIALTY: Computer engineering, Multimedia and Network.

DOCTORAL SCHOOL:
Ecole doctorale d'informatique, telecommunication et electronique de Paris. 46, rue Barrault 75014 Paris, FRANCE; http://www.telecom-paristech.fr/.

EXTERNAL LABORATORY:
University of Applied Sciences of Western Switzerland, Fribourg. 80, Bd de Perolles 1700 Fribourg, SWITZERLAND; http://www.eia-fr.ch.

THESIS ADVISOR:
Professor Ahmed Serhrouchni
46, rue Barrault 75014 Paris France
phone: +33 1 45 81 79 94; Ahmed.Serhrouchni@telecom-paristech.fr.

THESIS CO-ADVISORS:
Professor Omar Abou Khaled
Bd de Perolles 80, 1700 Fribourg Switzerland
phone: +41 26 429 65 89; Omar.AbouKhaled@hefr.ch
http://omar.aboukhaled.home.hefr.ch.
Professor Elena Mugellini
Bd de Perolles 80, 1700 Fribourg Switzerland
phone: +41 26 429 68 70; Elena.Mugellini@hefr.ch
http://elena.mugellini.home.hefr.ch.

THESIS FINANCING: EIAFR and CERN.

Résumé

Au cours de la dernière décennie, nous avons assisté á une croissance exponentielle de documents numériques et de ressources multimédias, y compris une présence massive de ressources vidéo. Les vidéos sont devenu de plus en plus populaire grâce au contenue riche á l'audio riche qu'elles véhiculent (contenu audiovisuelle et textuelle). Les dernières avancées technologiques ont rendu disponibles aux utilisateurs cette grande quantité de ressources multimédias et cela dans une variété de domaines, y compris les domaines académiques et scientifiques. Toutefois, sans techniques adéquates se basant sur le contenu des multimédia, cette masse de donnée précieuse est difficilement accessible et demeure en vigueur inutilisable.

Cette thèse explore les approches sémantiques pour la gestion ainsi que la navigation et la visualisation des ressources multimédias générées par les conférences scientifiques. Un écart, que l'on appelle sémantique, existe entre la représentation des connaissances explicites requis par les utilisateurs qui cherchent des ressources multimédias et la connaissance implicite véhiculée le long du cycle de vie d'une conférence. Le but de ce travail est de fournir aux utilisateurs une plateforme qui améliore la recherche de l'information multimédia des conférences en diminuant cette distance sémantique. L'objectif de cette thèse est de fournir une nouvelle approche pour le contenu multimédia basé sur la recherche d'information dans le domaine des conférences scientifiques.

La contribution de cette thèse se présente en trois parties. La première contribution est la conception et la mise en place de HELO, un modèle de connaissance ontologique que les modèles les informations véhiculées au cours du cycle de vie d'une conférence scientifique tels que les enregistrements vidéo, les fichiers de présentation, ou aussi l'information administrative. La deuxième contribution est la conception d'une approche intégrée basée sur HELO pour la gestion de l'information, rendant ainsi possible l'exploration de l'information générée au cours du cycle de vie d'une conférence. La troisième contribution est la conception d'une navigation ainsi qu'une approche de visualisations basées sur HELO permettant aux utilisateurs d'effectuer facilement leurs requêtes. Afin de valider l'approche proposée, nous avons mis en place CALIMERA, une plateforme intégrée pour la

recherche d'information multimédia appliquée au domaine des conférences scientifiques. CALIMERA offre un ensemble d'outils de gestion et de récupération. Elle est guidée par le modèle sémantique HELO et déployée sur un réseau pair á pair. La plateforme expérimentale proposée a été testé sur un ensemble de conférences qui ont eu lieu á l'EIAFR et au CERN.

Plateforme de recherche basée dur l'ontologie pour des données multimedia dans un réseau pair à pair

Maria SOKHN

Résumé

Au cours de la dernière décennie, nous avons assisté à une croissance exponentielle de documents numériques et de ressources multimédias, y compris une présence massive de ressources vidéo. Les vidéos sont devenu de plus en plus populaire grâce au contenue riche à l'audio riche qu'elles véhiculent (contenu audiovisuelle et textuelle). Les dernières avancées technologiques ont rendu disponibles aux utilisateurs cette grande quantité de ressources multimédias et cela dans une variété de domaines, y compris les domaines académiques et scientifiques. Toutefois, sans techniques adéquates se basant sur le contenu des multimédia, cette masse de donnée précieuse est difficilement accessible et demeure en vigueur inutilisable. Cette thèse explore les approches sémantiques pour la gestion ainsi que la navigation et la visualisation des ressources multimédias générées par les conférences scientifiques. Un écart, que l'on appelle sémantique, existe entre la représentation des connaissances explicites requis par les utilisateurs qui cherchent des ressources multimédias et la connaissance implicite véhiculée le long du cycle de vie d'une conférence. Le but de ce travail est de fournir aux utilisateurs une plateforme qui améliore la recherche de l'information multimédia des conférences en diminuant cette distance sémantique. L'objectif de cette thèse est de fournir une nouvelle approche pour le contenu multimédia basé sur la recherche d'information dans le domaine des conférences scientifiques. La contribution de cette thèse se présente en trois parties. La première contribution est la conception et la mise en place de HELO, un modèle de connaissance ontologique que les modèles les informations véhiculées au cours du cycle de vie d'une conférence scientifique tels que les enregistrements vidéo, les fichiers de présentation, ou aussi l'information administrative. La deuxième contribution est la conception d'une approche intégrée basée sur HELO pour la gestion de l'information, rendant ainsi possible l'exploration de l'information générée au cours du cycle de vie d'une conférence. La troisième contribution est la conception d'une navigation ainsi qu'une approche de visualisations basées sur HELO permettant aux utilisateurs d'effectuer facilement leurs requêtes. Afin de valider l'approche proposée, nous avons mis en place CALIMERA, une plateforme intégrée pour la recherche d'information multimédia appliquée au domaine des conférences scientifiques. CALIMERA offre un ensemble d'outils de gestion et de récupération. Elle est guidée par le modèle sémantique HELO et déployée sur un réseau pair à pair. La plateforme expérimentale proposée a été testé sur un ensemble de conférences qui ont eu lieu à l'EIAFR et au CERN.

La Figure 1 - Plan du contenu présente la structure de la thèse. Après une introduction générale du contexte, l'état de l'art est présenté. Les partie trois

et quatre sont consacrés à la contribution de la thèse ainsi qu'au prototype. La thèse est conclue par une synthèse et les perspectives futures.

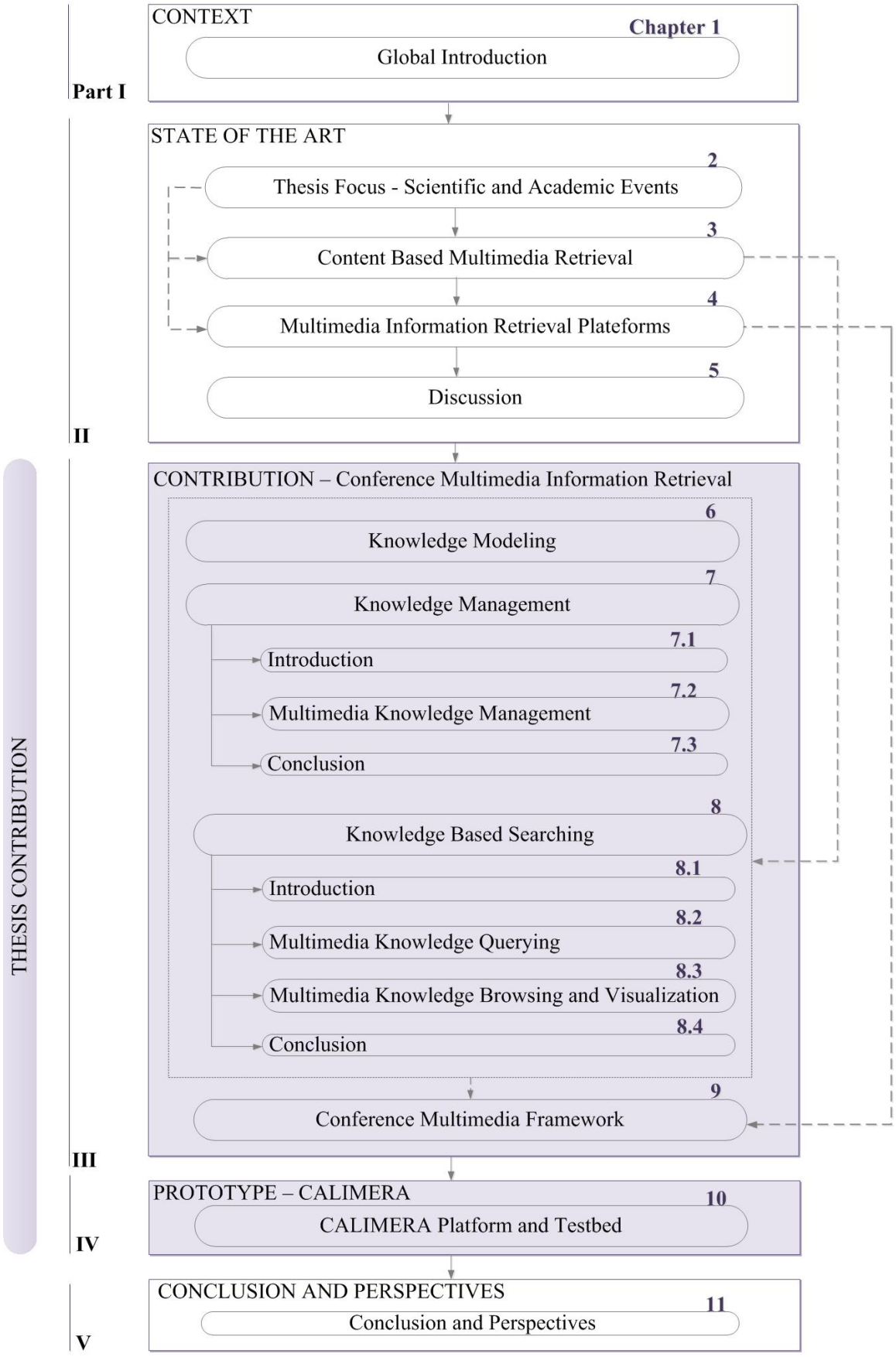


Figure 1 - Plan du contenu

Les progrès technologiques récents ont fait une grande quantité de ressources multimédia disponibles aux utilisateurs dans une variété de domaines. Toutefois, sans suffisamment des approches pour un accès effectif à des contenus multimédias, cette mine de données est difficilement accessible et, en effet inutilisable.

Dans cette thèse nous avons cherché à fournir une nouvelle approche pour la recherche par contenu de données multimédia pour les événements scientifiques et universitaires. Les informations diffusées par l'intermédiaire des conférences est en croissance constante et un grand nombre de ces événements sont enregistrés pour une utilisation ultérieure (lecture, recherche, préparation de cours, et ainsi de suite). De nombreuses caractéristiques propres aux manifestations scientifiques peuvent être identifiées et manipulées dans le but d'améliorer la recherche par contenu (cf. Figure 2 - Exemple d'information véhiculée dans un cycle de vie d'une conférence).

L'analyse globale de l'état de l'art nous a conduits à conclure à un ensemble d'observations et réflexions qui ont motivés les principaux axes de recherche du travail proposé dans cette thèse (cf. Figure 3 - Récapitulatif sommaire de l'état de l'art).

Dans cette thèse nous avons cherché à fournir une nouvelle approche pour le contenu multimédia basé sur la récupération. Nous avons exploré des approches sémantiques navigation de gestion de contenu basée sur la visualisation et des ressources multimédias généré pour et lors des conférences scientifiques. Nous avons cherché à fournir aux utilisateurs une plateforme qui améliore la recherche d'information de conférence en réduisant ce qui est communément nommé dans la littérature le «fossé sémantique». L'écart qui existe entre la représentation des connaissances explicites requis par les utilisateurs qui cherchent les ressources multimédia et les connaissances implicites véhiculées au sein du cycle de vie d'une conférence.

Trois questions principales ont été abordées et la principale contribution a été portée vers ces questions :

La première question est la modélisation des connaissances (cf. Figure 4 - HELO, représentation d'une partie des concepts): Nous avons conçu un modèle ontologique qui a été destiné à être utilisé comme un modèle sous-jacent pour le processus de recherche d'information multimédia (cf. Figure 6 - CALIMERA, représentation en couches, Figure 7 - CALIMERA, Cycle de vie d'une conférence).

L'état de l'art a montré l'étude qu'une attention croissante est accordée à l'approche basée sur l'ontologie, car il offre des solutions prometteuses pour améliorer et permettre la recherche d'informations basée sur le contenu. Les modèles existants couvrent une partie des structures et le contenu requis descriptions basées dans le domaine des conférences scientifiques (cf. Figure

4 - HELO, représentation d'une partie des concepts). Nous avons conçu et mise en œuvre HELO, une ontologie qui modélise les informations transmises au cours du cycle de vie d'une conférence scientifique. Le modèle est composé de concepts granulés qui sont représentatifs de l'annotation des utilisateurs, et de huit différents concepts que nous appelons des *Scope* qui sont représentatifs de la navigation des utilisateurs et du processus de visualisation (cf. Figure 5 - HELO, représentation).

La deuxième question abordée est la gestion des connaissances: Dans cette thèse nous avons cherché à améliorer la description du contenu multimédia en se basant sur l'agrégation de l'information disponible générée par le cycle de vie d'une conférence (cf. Figure 8 - CALIMERA, Cycle de vie d'une conférence). Pour résoudre ce problème, nous avons conçu une approche intégrée pour la recherche d'information par contenu en se basant sur la modélisation du domaine de conférences scientifiques. L'approche est basée l'ontologie de conférences. Nous avons cherché à rendre possible l'utilisation de l'information existante transmises au cours un cycle de vie d'une conférence. Cela a été possible en intégrant cette conférence dans le modèle ontologique HELO (cf. Figure 8 - CALIMERA, Cycle de vie d'une conférence).

La troisième question abordée est la navigation et la visualisation des connaissances: Nous avons cherché à améliorer la récupération multimédia en proposant un modèle approche fondée sur la navigation. Nous avons proposé une approche de navigation et de visualisation sur la base de HELO qui prend en compte le cycle de vie de récupération d'informations.
Nous avons cherché à donner aux utilisateurs la possibilité d'explorer l'information sur la conférence par rapport aux *Scope* définies sur la base de leurs exigences. Afin d'unifier les nombreux efforts qui ont été menés au cours de la thèse nous avons intégrés les approches proposées dans la plateforme de recherche d'information. Nous avons conçu une plateforme intégrée pour la recherche d'information multimédia basée sur le contenu (cf. Figure 6 - CALIMERA, Architecture). Cette approche, fait usage du cycle de vie grâce au modèle sémantique HELO sous-jacent à l'architecture.

Nous avons cherché dans cette démarche de thèse à faire usage du potentiel du P2P, principalement pour l'annotation de contenu et la recherche.

Pour valider notre approche, nous avons mise en place CALIMERA (cf. Figure 6 - CALIMERA, Architecture, Figure 7 - CALIMERA, représentation en couches), une plateforme de recherche d'information guidée par une ontologie qui intègre un ensemble d'outils pour la récupération des ressources de conférence multimédia (cf. Figure 8 - CALIMERA, Cycle de vie d'une conférence).

Au niveau de l'annotation, le travail de thèse présente une amélioration significative dans l'annotation d'enregistrement vidéo-conférence: (1) Nous avons fait usage de l'information existante qui n'a pas été utilisé dans les approches précédentes. (2) Nous avons conçu et mis en œuvre un algorithme de segmentation de vidéo-conférence. L'algorithme proposé repose sur la détection de changement. Cette décision était fondée sur le fait que les utilisateurs peuvent avoir besoin de visualiser une séquence vidéo spécifique d'une conférence. Cette séquence la plupart des cas correspond à une diapositive spécifique ou un ensemble de diapositives. (3) Nous avons proposé une annotation automatique par le biais un algorithme de speech-to-text qui attribue automatiquement l'annotation basé sur le temps. (4) Nous avons offert aux utilisateurs un outil perfectionné, conçu pour l'annotation vidéo conférence. Cet outil d'annotation semi manuel est basé sur HELO un modèle d'ontologie commune (cf. Figure 9 - CALISEMA, outil d'annotation basée sur HELO).

Au niveau de navigation et de visualisation, le travail de thèse présente une nouvelle approche de la recherche et la visualisation des conférences basées sur une ontologie de la conférence. (1) Nous offrons aux utilisateurs la possibilité de naviguer à travers les ressources multimédias basé sur le modèle sémantique. Grâce à cette approche de navigation, fait référence dans cette thèse travailler comme la recherche à facettes, a permis aux utilisateurs d'explorer les informations disponibles et de visualiser le résultat de la recherche en fonction de différents champs d'application tels que PersonScope, CommunityScope, TemporalScope, et ainsi de suite. (2) Nous avons également offert aux utilisateurs la possibilité de filtrer les résultats en fonction de plus d'un champ d'application. Cette approche est essentielle et intéressante en raison de filtrage des données des aspects qui, selon les cas, peut en outre améliorer la visualisation du résultat de recherche (Figure 10 - SMAC, outil de visualisation basée sur HELO).

Enfin, CALIMERA présente un cadre de bout en bout intégrée, qui peut être utilisé comme une base utile pour les expériences, principalement pour une plate-forme opérationnelle mais plus probablement comme une première étape pour l'évaluation des utilisateurs. Les résultats actuels obtenus dans ce travail de thèse appel pour diverses directions de recherche futures. Dans la section suivante, nous proposons quelques-unes de ces améliorations et possibilités d'extension.

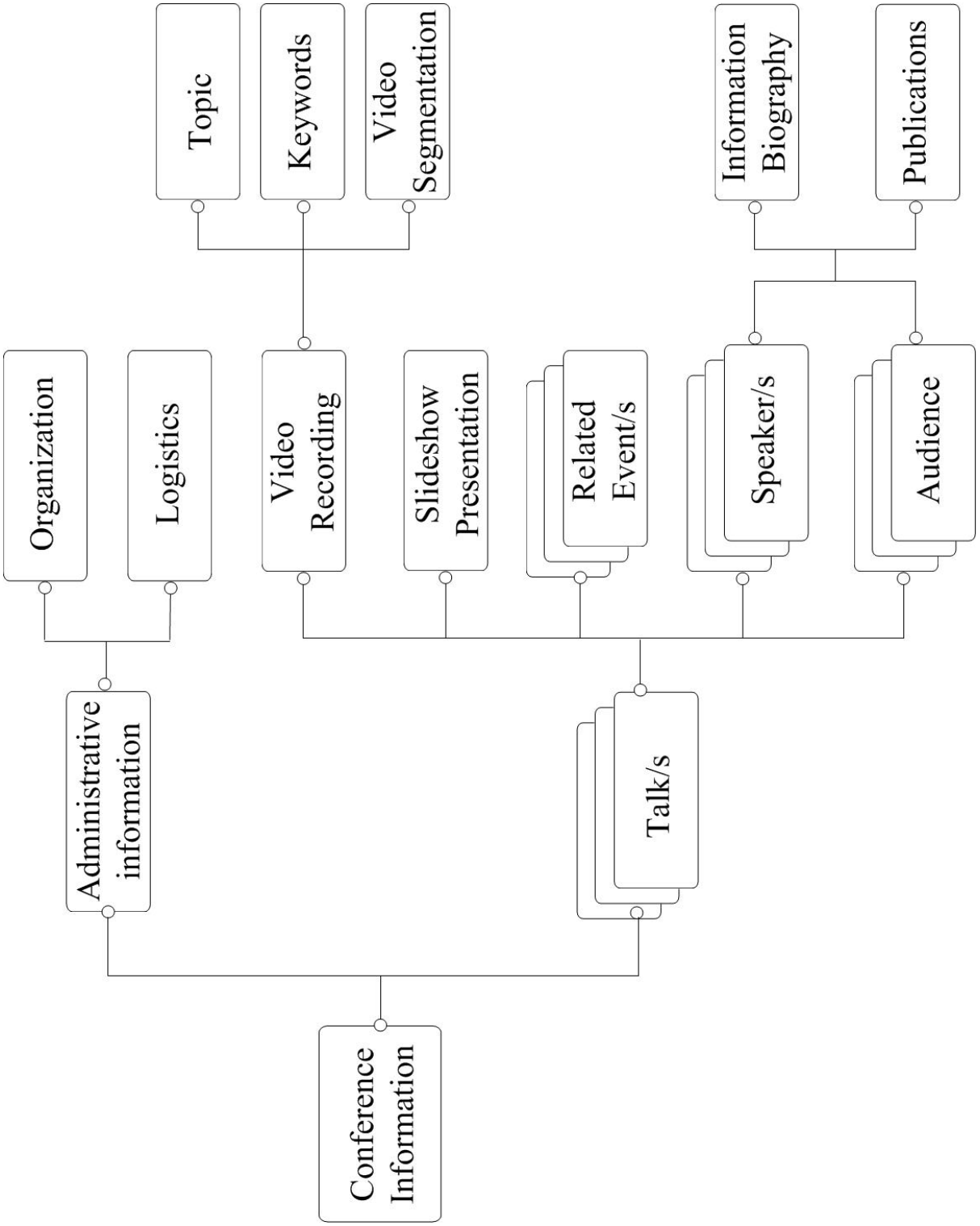


Figure 2 - Exemple d'information véhiculée dans un cycle de vie d'une conférence

Figure 3 - Récapitulatif sommaire de l'état de l'art

Issues	Sub-issues	ISWC [Knud Moller 2007]	VIDEOLECTURES [VID b]	VIKEF [VIK]	MATTERHORN [MAT]	SAPIR [SAP]	RKBEXPLORER [Glaser 2007]	SEMANTICLIFE [Hoang 2006]	DBLPVIZ [DBL a]	ARNETMINER [ARN]	**THESIS CONTRIBUTION** CALIMERA [Sokhn 2010b, Sokhn 2011b]
Models Designing	Conference Domain	■	■	■	■		■		■	■	■
	Conference Knowledge Modeling	■					■				■
	Video Conference Ontology Model										■
Multimedia Management	Conference Life Cycle Knowledge Management				■						■
	Ontology Driven Knowledge Management			■			■				■
	Multimedia Synchronization				■						■
	Hybrid Knowledge Management				■						■
Multimedia Browsing and Visualization	Multimedia Browsing and Visualization		■	■		■	■			■	■
	Knowledge Driven Browsing and Visualization			■							■
Integrated Framework	Tools Independent Framework				■						■
	Usage of P2P Potential for Multimedia Retrieval					■					■
	Multi-format Metadata							■			■

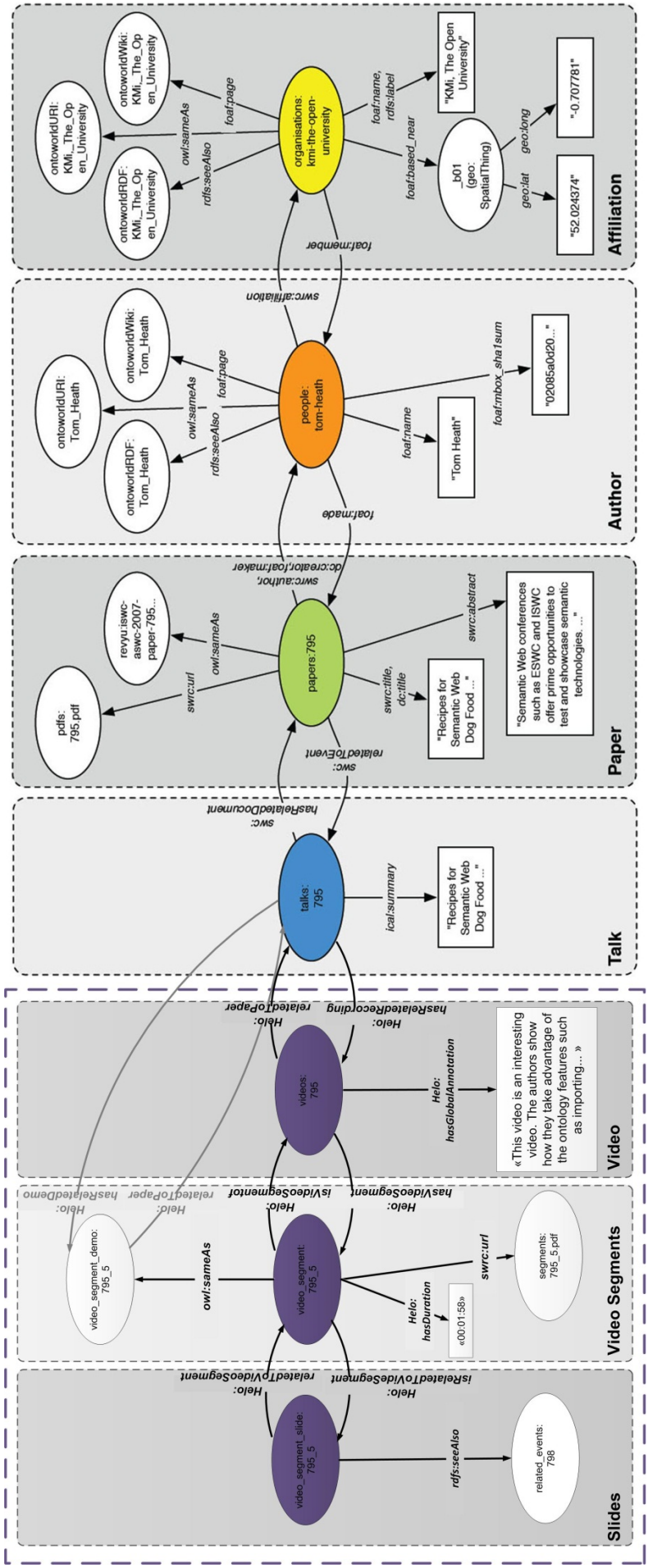


Figure 4 - HELO, représentation d'une partie des concepts

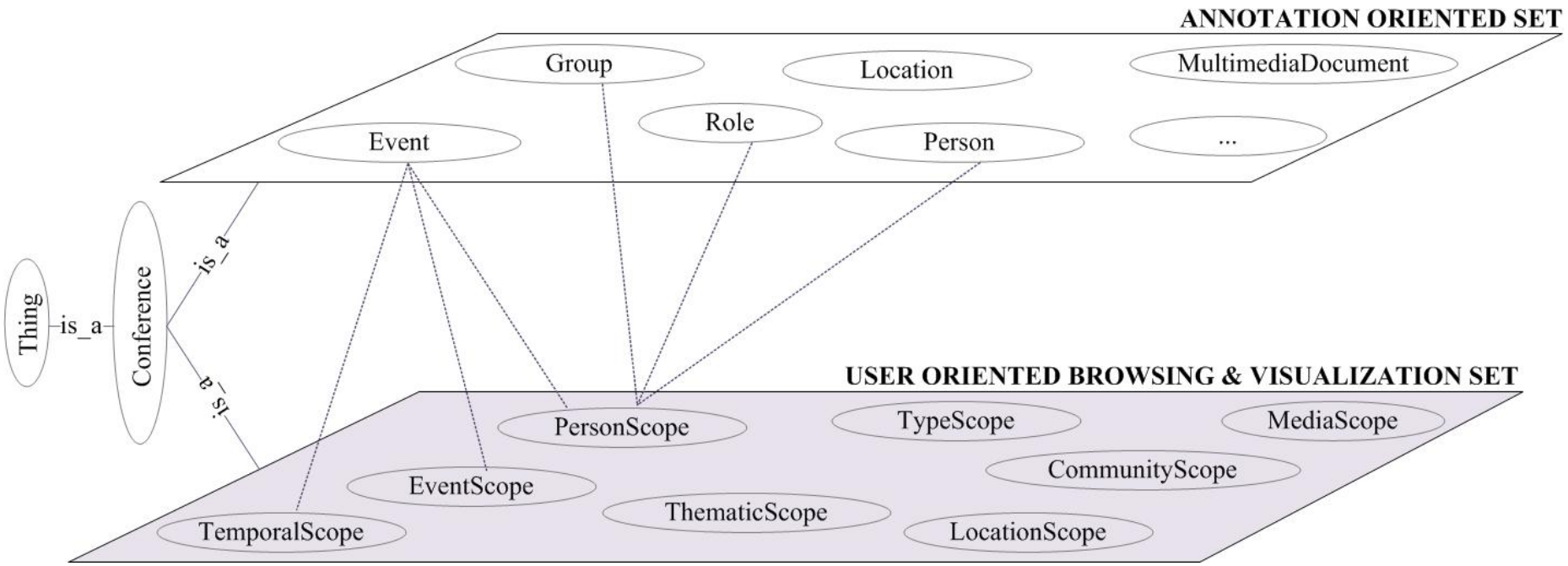


Figure 5 - HELO, représentation

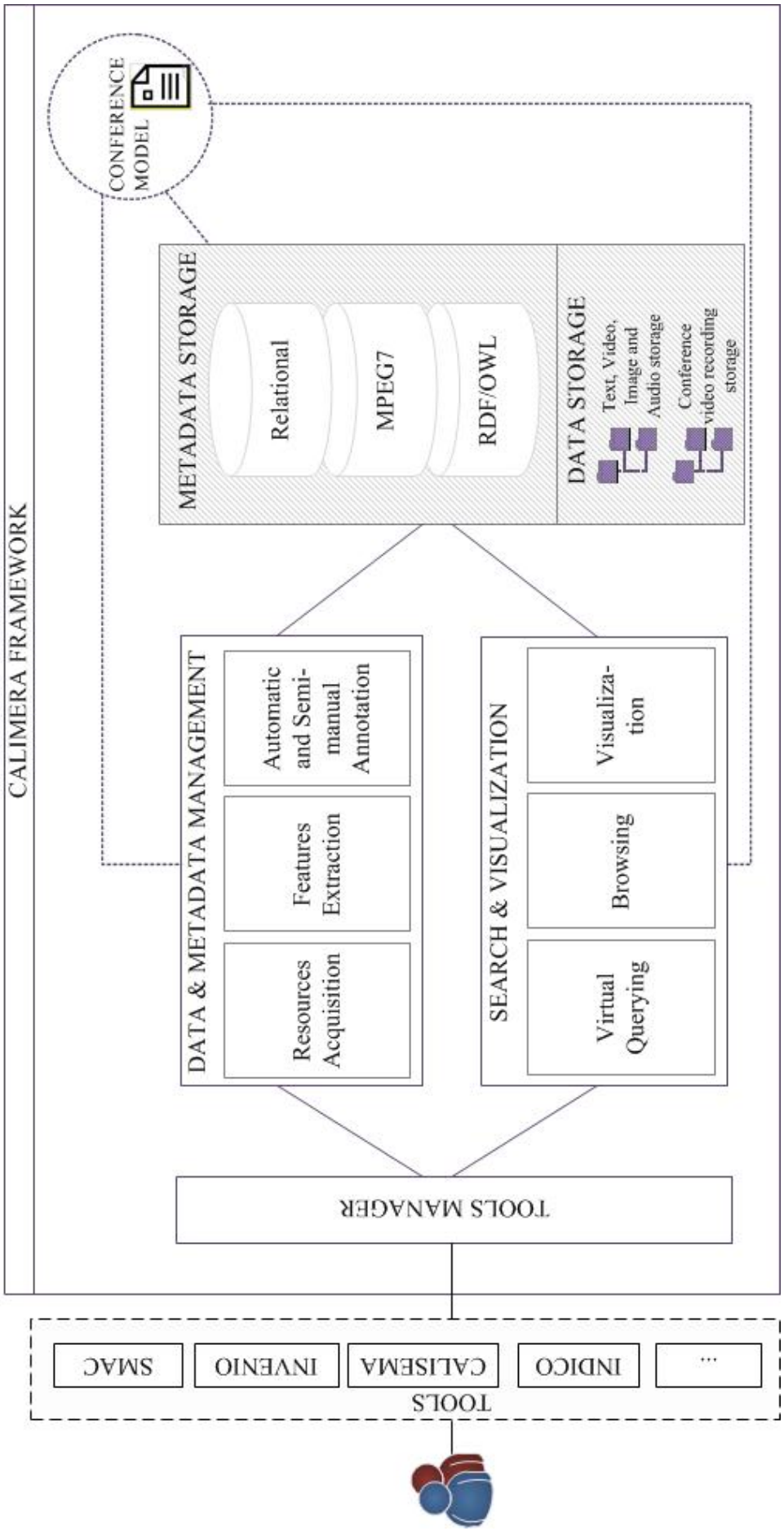


Figure 6 - CALIMERA, Architecture

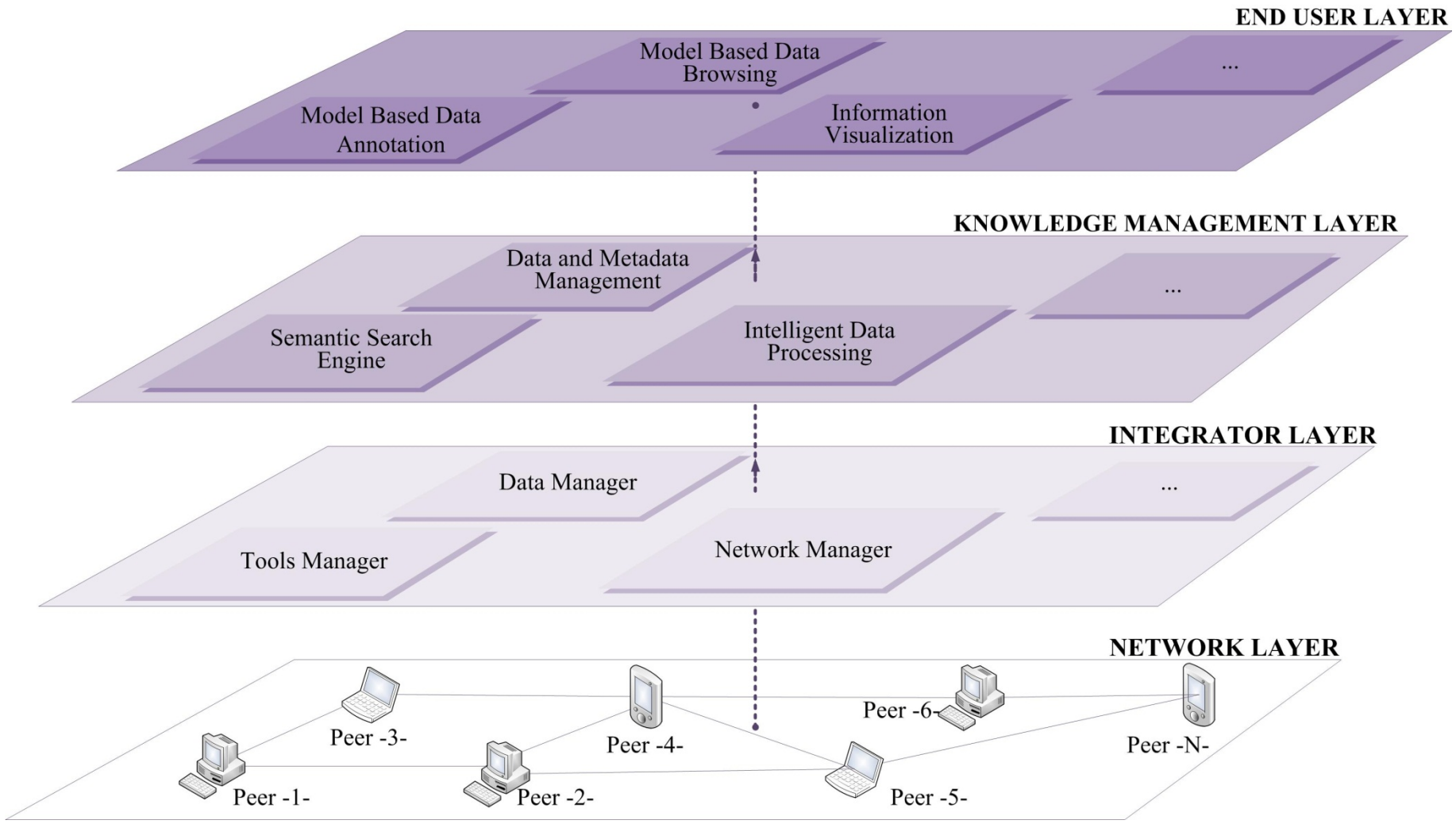


Figure 7 - CALIMERA, représentation en couches

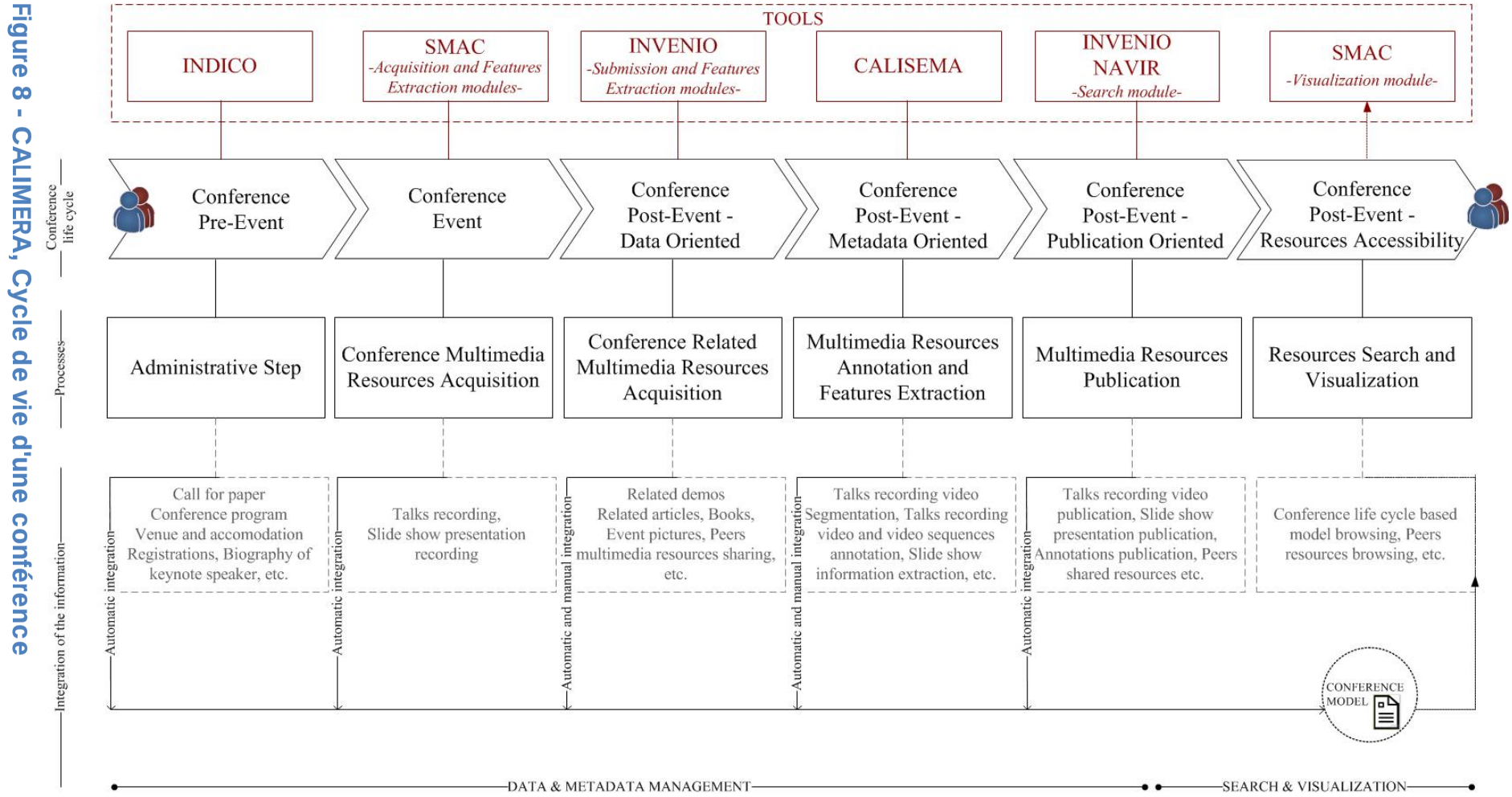


Figure 8 - CALIMERA, Cycle de vie d'une conférence

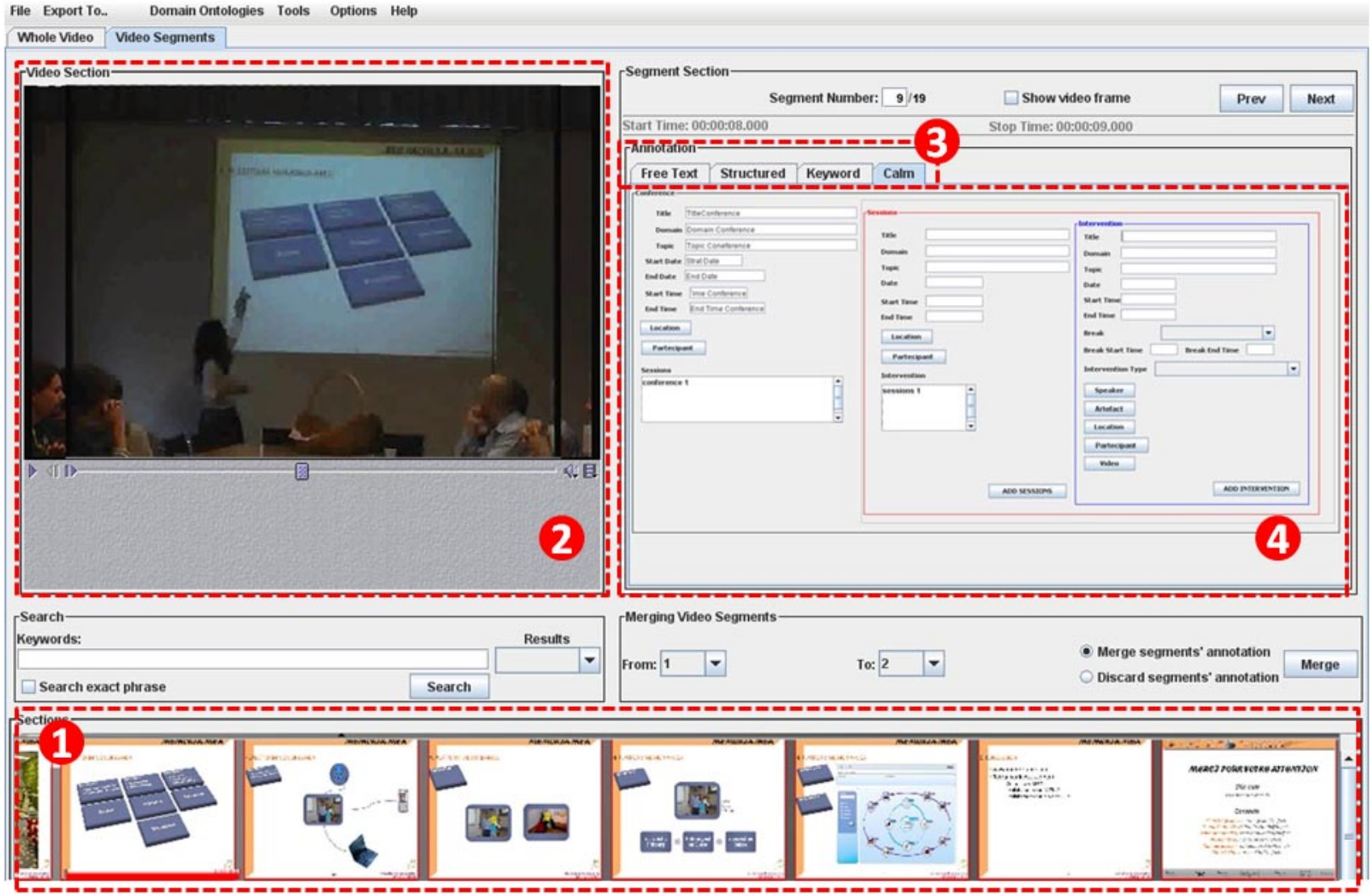


Figure 9 - CALISEMA, outil d'annotation basée sur HELO

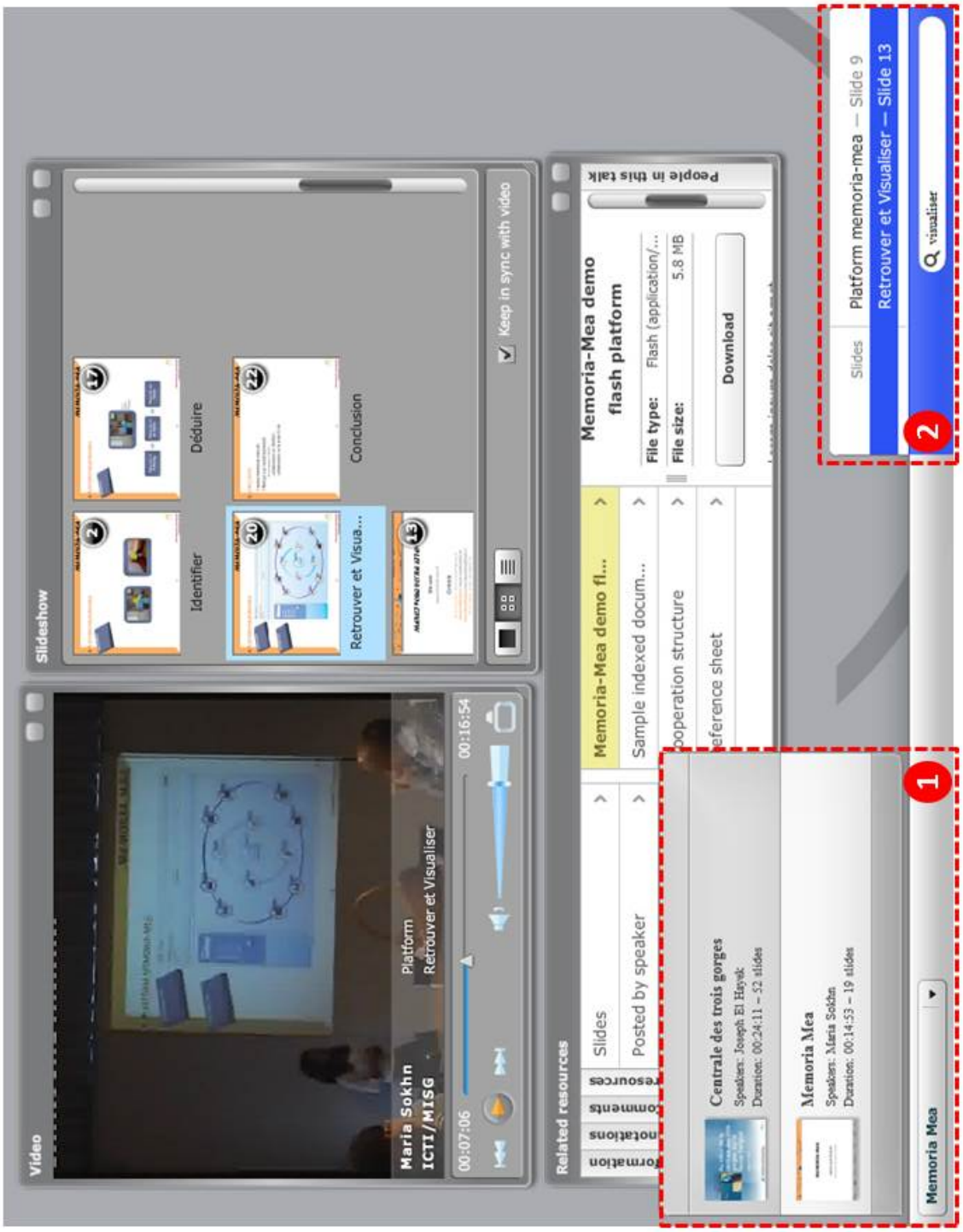


Figure 10 - SMAC, outil de visualisation basée sur HELO

Pour la suite, une des tâches imminente à faire est l'évaluation de la plateforme proposée en particulier lorsque les recherches sont effectuées sur le réseau P2P à grande échelle. Comme mentionné précédemment, le succès d'un système de recherche multimédia est principalement déterminé par sa capacité à soutenir les besoins des utilisateurs. À cette fin, nous avons décidé de concevoir la plateforme en se basant sur un domaine, le domaine spécifique. Cette spécification nous permet de se concentrer sur besoins spécifiques des utilisateurs. L'évaluation des différentes caractéristiques intégrées dans le cadre à savoir le modèle, les interfaces utilisateur graphiques à la fois pour l'annotation et recherche sont d'un grand intérêt pour améliorer l'approche proposée et par conséquent la plateforme CALIMERA.

Cette évaluation devrait exploiter l'évaluation de la ontologique modèle HELO, l'évaluation de l'interface sémantique d'annotation, l'évaluation de l'interface de moteur de recherche sémantique, et l'évaluation de la rendu multimédia et interfaces de navigation. Comme prochaines étapes, il serait d'intérêt pour améliorer l'approche proposée. Cela pourrait se faire à différents niveaux: (1) Dans le domaine des données et des métadonnées de gestion, des améliorations peuvent être transportés sur le domaine de l'extraction de contenu. En effet, de nombreux niveau de fonctionnalités, telles que la reconnaissance du visage, qui peut être extraite peut potentiellement être utilisé pour l'indexation et la récupération. Techniques d'annotation sémantique comme reconnaissance de caractères vidéo optique (OCR), suivi des objets peut être aussi intégrés dans la plate-forme pour enrichir l'annotation disponibles. (2) Dans le domaine de la recherche, l'approche requête virtuel doit être mis au point et à évaluer. Plus interfaces utilisateurs devraient soutenir l'insertion de données, mise à jour et la pertinence des commentaires. Les interfaces utilisateur proposées devraient être complétées par la mise en œuvre celles qui n'ont pas encore été développé. (3) Dans le domaine du réseau P2P, un rôle important et aspect intéressant peut être amené à améliorer la performance de la plate-forme. Comme la plateforme proposée est modulaire, nous pouvons répartir les différents modules au cours de la connecté pairs faisant usage de leur unité de traitement de commande (CPU). Cette distribution est intéressante pour le module vocal-texte et l'outil d'enregistrement principalement lors de des séances parallèles qui doivent être enregistrées.

D'autres aspects importants pourraient être prises en compte dans ce domaine de travail et peut être considérée comme une extension de la plateforme proposée. Il peut s'agir de l'extension de la HELO modèle, y compris d'autres ontologies existantes et de faire l'utilisation de la grande quantité d'informations que l'on peut trouver sur le web. Informations tels que les bibliothèques numériques de la littérature scientifique et d'information que nous pouvons extraire à partir de réseaux sociaux actuels et des réseaux P2P.

Abstract

During the last decade we have witnessed an exponential growth of digital documents and multimedia resources, including a vast amount of video resources. Videos are becoming one of the most popular media thanks to the rich audio, visual and textual content they may convey. The recent technological advances have made this large amount of multimedia resources available to users in a variety of areas, including the academic and scientific realms. However, without adequate techniques for effective content based multimedia retrieval, this large and valuable body of data is barely accessible and remains in effect unusable.

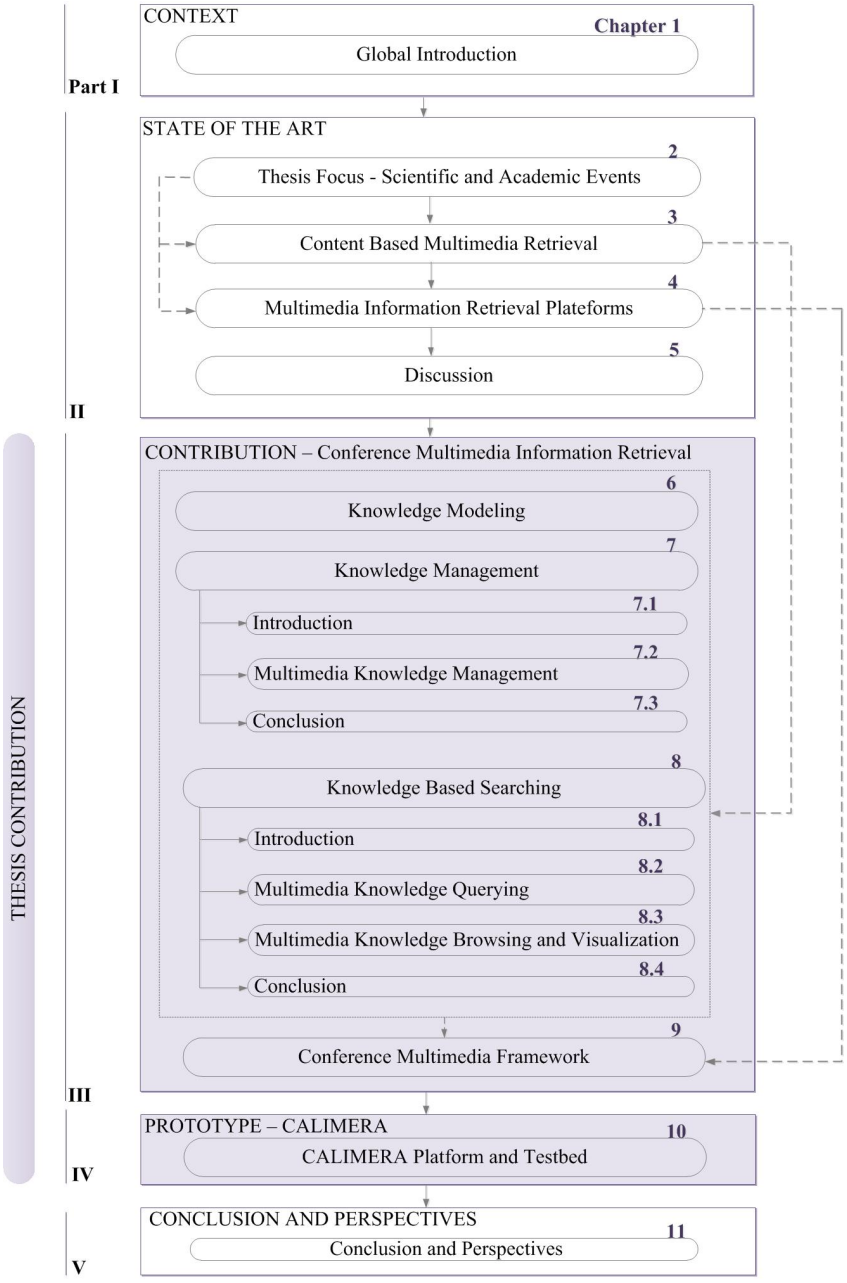
This thesis explores semantic approaches to content based management browsing and visualization of the multimedia resources generated for and during scientific conferences. Indeed, a so-called semantic gap exists between the explicit knowledge representation required by users who search the multimedia resources and the implicit knowledge conveyed within a conference life cycle. The aim of this work is to provide users with an integrated framework that enhances the conference multimedia information retrieval by bridging this semantic gap. The goal of this thesis is to provide a novel approach for content based multimedia retrieval in the domain of scientific conferences.

The contribution of this thesis is of threefold. The first contribution is the design and implementation of HELO, an ontological knowledge model that models the information conveyed during the life cycle of a scientific conference such as video recordings, talk presentation files, speaker information and the administrative information. The second contribution is the design of an integrated approach for knowledge based management guided by HELO, making therefore use of the existing unused information conveyed during a conference life cycle. The third contribution is the design of a knowledge based browsing and visualization approach enabling users to easily make complex and structured queries. In order to validate the proposed approach, we put forward CALIMERA, an integrated framework for conference multimedia retrieval. CALIMERA offers a set of management and retrieval tools. It is driven by the semantic model HELO and deployed over a Peer to Peer network. The proposed experimental framework

was tested over a set of conferences that took place at the EIAFR and the CERN.

Table of Contents

Contents Road Map



List of Figures

WRITING CONVENTIONS

- Titles and important words and sentences are emphasized in **BOLD**
- New concepts/words are written in *ITALIC*
- Cited sentences and words are "QUOTED"
- Query examples are "*QUOTED AND IN ITALIC*"
- Bibliographic references are cited [BETWEEN BRACKETS IN RED]
- Acronyms and a list of particular terms are provided in the APPENDIX B

Part I

CONTEXT

Chapter 1

Global Introduction

1.1 Background and Motivation

During the last decade we witnessed an exponential growth of digital documents and multimedia resources, including a vast amount of video resources. Videos are becoming one of the most popular media thanks to the semantic audio, visual and textual content they convey. The recent technological advances have made this large amount of multimedia resources available to users in a variety of areas, including the academic and scientific realms. However, without adequate techniques for effective access to multimedia content, this large and valuable body of data is barely accessible and in effect unusable. To date, most effort has been directed toward systems for image and audio retrieval as these media were in use much earlier than video. The multimedia retrieval systems that focus on video resources therefore require adapted and novel techniques and approaches for providing relevant and efficient content based multimedia management.

However, content based multimedia retrieval systems are difficult to build, mainly because of the large difference between the content that can be automatically extracted and the content that needs to be extracted in order to satisfy the needs of the users during the retrieval process. This is especially the case in the academic and the scientific domains, where the content that can be automatically extracted consists in content such as the shape, color, motion, and the like and the content that users who are searching the multimedia resources require, consists in content such as important topics, events, relation to other events and community. In the information retrieval field, this gap is called, the "**semantic gap**". The "**semantic gap**" has been defined and referenced in several research works [Liu 2007, Hellmann 2010, Ma 2010]. Smeulders et al. [Smeulders 2000] define it as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation".

The main challenge facing multimedia retrieval systems today is bridging this "**semantic gap**". Several approaches have been proposed. One of these approaches is the domain specific approach which exploits the typical characteristics of a particular video genre to design the most effective tools for information management (extraction, annotation, etc.) and retrieval (querying and browsing). In this thesis we focus on the scientific conference domain. We define the word

'conference' in this thesis is as follows: A conference, is any type of meeting, happening at a determinable time and place, with at least one speaker who intend to inform, explain, or talk about one or several topics. As technological developments have made recording technology more affordable and easier to use, the number of conference presentations that are recorded has been increasing. However, the information conveyed within the resulting multimedia resources is hardly accessible by large numbers of various interested users, including professors, students and researchers. In addition to the stated context, the characteristics of conferences make the use of the potential of the Peer to Peer (P2P) a promising solution for the enhancement of the multimedia information retrieval. Indeed, among several definitions found on the web, a conference is considered as a large gathering of individuals or members of one or several organizations, for discussing matters of common interest [COF a], an occurrence happening at a determinable time and place, with or without the participation of human agents [COF b], a formal meeting that typically takes place over a number of days and involves people with a shared interest [COF c]. In this context, the resources are obviously distributed, since it is a gathering of individuals and members. This recalls also to the necessity of sharing the distributed information within the community. The large scale of the scientific community and its computational resources can be exploited by using the P2P potential, hence improving the multimedia retrieval process [Ramzan 2010].

1.2 Thesis Objectives

The goal of this thesis is to provide a novel and integrated[1] approach for content based multimedia retrieval in the domain of scientific conferences. The aim of this thesis is to provide users with an integrated framework that enhances the conference multimedia information retrieval by bridging the so-called "**semantic gap**". We aim to take advantage of the information conveyed during the life cycle of a conference. This information includes (Fig. 1.1): The administrative information (e.g. conference organization, logistics), the slideshow presentation of the talk (e.g. PowerPoint or PDF formats), the video recording of each talk with its content (e.g. video segmentation, keywords, topics), the related events, the speaker and the audience (each characterized by such information as name, organization the person belongs to and publications), the related demos, and more. The ability to extract this information constitutes a significant advance in the multimedia information retrieval, since it enriches its content based description. To that end, this thesis explores semantic approaches to content based management browsing and visualization of the multimedia resources generated for and during scientific conferences. It also explores the potential of P2P networks for supporting existing approaches to multimedia information retrieval. The main achievements of the thesis are presented

[1] In this work, the word 'integrated' is used to refer to an effective unit that combines different aspects/approaches that are relevant to address a single issue, which is, in this case, the enhancement of the content based multimedia retrieval

in the following section.

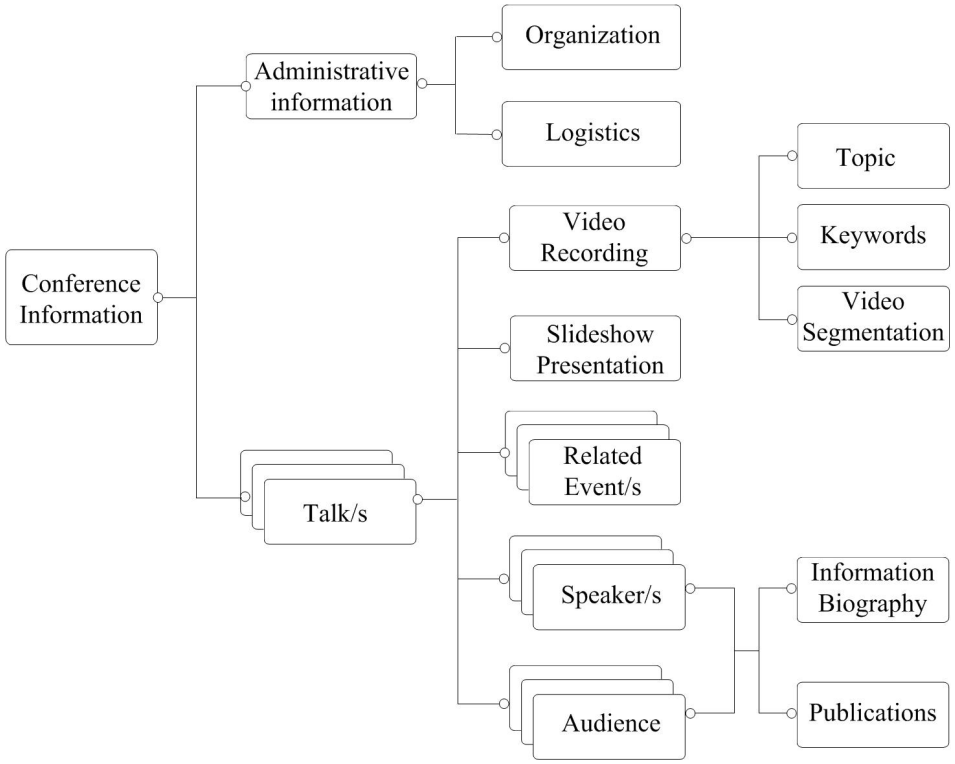


Figure 1.1: Example of the information conveyed within a conference life cycle

1.3 Summary of Achievements

In this thesis we present the design and implementation of an integrated framework for content based multimedia retrieval that is driven by a knowledge[2] model and deployed over a P2P network. We also offer a set of tools to validate the proposed approach. The main results of this thesis can be summarized as follows:

- **Conference knowledge modeling:** The design and implementation of an ontological model that formalizes the information conveyed during the life cycle of a scientific conference such as video recordings, talk slideshow presentation files, speaker information and administrative information. The designed model is named HELO which stands for High-level modEL Ontology. HELO is composed of granular concepts that are representative of users' annotation

[2] According to Oxford Dictionaries Online the term 'knowledge' is "Facts, information, and skills acquired through experience or education; the theoretical or practical understanding of a subject'. In this thesis, the term 'knowledge' is used to designate an understanding of a specific domain that goes beyond the simple information and integrates aspects such as competences, skills, expertise, etc.

process and of eight different *scopes*[3] that are representative of users' browsing and visualization processes [Sokhn 2010b, Sokhn 2009a, Sokhn 2009d].

- **Hybrid knowledge management:** The design of an integrated approach for content based information retrieval in the domain of scientific events. The approach is based on the conference ontology HELO. It combines the automatic, manual and semi-manual techniques. Thanks to the unified model the proposed approach makes use of the existing unused information conveyed during a conference life cycle to enrich the multimedia content description [Sokhn 2011a, Atitallah 2011a, Atitallah 2011b, Revertera 2011, Sokhn 2010a, Sokhn 2009b].

- **Knowledge browsing and visualization:** The design of a model based browsing and visualization approach based on the semantic conference model. This approach gives users the ability to explore conference information with respect to *scopes* defined based on their requirements. Because the multimedia content description is based on a model that structures characteristics and relationships through a set of *scopes*, this enables users to easily make complex and structured queries [Sokhn 2010b, Sokhn 2011b, Dumoulin 2011, Dumoulin 2010, Sokhn 2009c].

- **Conference multimedia framework:** The implementation of the Conference Advanced Level Information ManagemEnt & RetrievaAl (CALIMERA) an integrated framework for the management and retrieval of conference multimedia resources. To validate the proposed approach we developed a set of tools for both the management and retrieval of conference multimedia resources. These tools, which are integrated into CALIMERA, are designed on top of the HELO ontological model [Sokhn 2010b, Sokhn 2011a, Sokhn 2009b].

1.4 Report Road Map

In this section we present an overview of the thesis's chapters organization.

Part I - CONTEXT

- **Chapter 1 - Global Introduction**
  In this introductory chapter we present the global context of this work, along with the thesis objectives and a summary of the main achievements.

Part II - STATE OF THE ART

- **Chapter 2 - Thesis focus - Scientific and Academic Events**
  In this chapter we present the background and the motivation for the thesis.

[3] In this thesis, the term '*scope*' designates a set of ontological concepts that refers to different views of a conference. This concept is detailed in the chapter 6.

- **Chapter 3 - Content Based Multimedia Retrieval**
  In this chapter we present an overview of the state of the art in content based multimedia retrieval. We focus on three main issues: The multimedia information modeling, the multimedia information management (content based acquisition, extraction, integration and annotation), and the multimedia information searching (querying and browsing). In the first part we present the semantic models and the different techniques that are currently used to semantically model specific domains, particularly the conference domain. We focus on the existing model that is used for the domain of scientific conferences. In the second part we present an overview of existing projects and techniques for multimedia information management, with a focus on the content based annotation approach. In the third part we present an overview of previous work done in the area of content based searching (querying and browsing) and visualization.

- **Chapter 4 - Multimedia Information Retrieval Platforms**
  In this chapter we present an overview of the major current multimedia information retrieval frameworks and platforms focusing on the frameworks developed for scientific and academic contexts. We analyze the three frameworks that we consider to be most relevant ones according to this thesis topic. In this chapter we also introduce P2P networks, as we believe that these may enhance the information retrieval process. Indeed, conferences being a gathering of individuals and members, this contributes to the existence of distributed information and the need of sharing and accessing resources and knowledge within a community. The large scale of the scientific community and its computational resources can be exploited by using the P2P potential.

- **Chapter 5 - Discussion**
  In this chapter we discuss the approaches and techniques presented in the previous chapters. We present the issues that have been identified and indicate the issues that we will address in this work. We discuss content based browsing and visualization approaches in detail and briefly discuss content based querying over heterogeneous data. The discussion in this chapter motivates the ontology driven framework approach for multimedia information retrieval in P2P networks.

Part III - CONTRIBUTION - Conference Multimedia Information Retrieval

- **Chapter 6 - Knowledge Modeling**
  In this chapter we present HELO, an ontology model for conferences. This model allows us to effectively manage and reuse knowledge conveyed in the content multimedia resources during the life cycle of a conference, specifically those resources used in scientific conferences. HELO is based on existing ontologies, integrating the content descriptions that specifically relate to scientific conference.

- **Chapter 7 - Knowledge Management**
  In this chapter we detail the proposed ontology driven knowledge management specifically features extraction and content annotation. In the first part we describe the proposed method for automatic metadata extraction and in the second part we describe the proposed semi-manual approach to ontology based annotation.

- **Chapter 8 - Knowledge Based Searching**
  In this chapter we describe the ontology driven knowledge searching and visualization. In the first part we introduce the proposed approach to heterogeneous data querying (a minor focus was given to this specific issue), and in the second part we describe the information browsing and visualization approach which is based on the semantic model HELO.

- **Chapter 9 - Conference Multimedia Framework**
  In this chapter we present the framework approach that we have designed for conference multimedia retrieval. The framework is driven by the HELO ontology. We describe the different modules constituting the framework.

Part IV - PROTOTYPE - CALIMERA

- **Chapter 10 - CALIMERA Platform and Testbed**
  In this chapter we present CALIMERA, which is the experimental research project that we set up as a proof of concept of the approach that we propose in this thesis. CALIMERA is implemented on a P2P network and integrates a set of tools for both information management and information retrieval. The design of these tools is based on by the semantic model HELO. In order to validate the thesis experimental project we set up a testbed of several videos provided by the CERN and performed CALIMERA over different conferences that took place at the EIAFR, the most important of which was the international Atlantic Web Intelligence Conference 2011 (AWIC2011).

Part V - CONCLUSION AND PERSPECTIVES

- **Chapter 11 - Conclusion and Perspectives:**
  In this final chapter, we state the conclusions and propose relevant directions for future research based on the work presented in this thesis.

Part II

STATE OF THE ART

Chapter 2

Thesis Focus - Scientific and Academic Events

In this section we motivate the thesis focus on the realm of scientific and academic events[1] by identifying the main aspects and characteristics of these events.

Scientific and academic events are gradually increasing [NBC], and the quantity of high quality scientific information disseminated via conferences is growing steadily [Sidiropoulos 2005]. A large number of these events are being recorded for subsequent use (e.g. replay, research, course preparation). In the last decade, researchers have addressed the issue of enhancing the content based information retrieval. Various approaches and systems have been proposed for scientific and academic events. Most of these approaches, addressed a specific issue related to the content based multimedia retrieval. Due to the dynamic and temporal nature of video, the content based video retrieval issue was more particularly addressed.

Every conference has a life cycle divided into different steps. Each of these steps conveys a set of information that can be used in the information retrieval process. The retrieval of a talk video recording may be considerably enhanced if the information generated through the conference life cycle is captured. What characterizes the video of scientific conferences is the domain characteristics relative to academic and more precisely scientific realm. Video recordings of scientific conferences have their own specificities such as the structure of the video, the scientific community relationships, the relative procedures, the related events and so on. In a scientific conference we may be interested to replay the different talks of a specific session, of a specific team, of a specific topic or year. We may be interested to watch back a specific sequence of the video relative to a specific slide of the presentation or relative to the demo part. Many characteristics specific to the scientific events can be identified and handled in order to enhance the content based multimedia retrieval. Adding to this, the conference talks are often recorded in poor lightening condition and have no attached presentations' outline, which makes the video analysis and semantic features extraction a challenging task. Moreover, scientific conferences are characterized by the social/scientific community which actively participate in scientific events and can consequently provide interesting multimedia resources (photos, related demo, books, papers, and more) that can be

[1] The term 'events' denotes a scheduled activity. In this work, this term designates primarily conferences.

subsequently linked to the recorded talk.

The multimedia resources are stored and subsequently searched and accessed by different category of users:

- Occasional users which may have specific needs without having necessarily the adequate keywords or knowledge to formulate the query. For example a student searching for a specific video sequence which describes what an ontology is. It may be a talk of a famous speaker.
- Advised users, i.e. users that may need an advanced and scientific based research/browsing model. For example professors that may be preparing their presentations, may need different video sequences illustrating their ideas. It may also be some participants of the conference that would like to see back their presentation or search for the event related photos for instance.
- Expert users that need to analyze different talks, to learn more about a topics or even to organize a scientific event. For example advisors that need to evaluate their student' presentation comparing to other presentations or professors that are preparing their courses may seek for specific details of a chosen topic, research community, or laboratory.

Figure 2.1: CERN Document Server (CDS) - portal

In addition to the described context, the thesis focus on scientific and academic events is motivated by the collaboration with the CERN. Indeed, at CERN, over

800,000 bibliographic records and 350,000 fulltext documents, organized in more than 500 collections, articles, books, journals, videos and more, are available, and the number of resources is continuously growing (Fig. 2.1). In this context, we identify a strong need for a content based multimedia retrieval in the scientific realm. Thanks to this collaboration we can make use of existing tools developed at the CERN for the conference management namely INDICO and INVENIO both of which are presented in the chapter 10. Moreover, having access to a full-scale data set is a motivating aspect for an efficient evaluation process for the thesis proposed approach.

Based on this context we propose in this thesis an approach that takes into account the characteristics of the scientific domains in order to provide users with a novel approach for multimedia information retrieval. The following part presents the state of the art related to the thesis focus and motivates the thesis approach for an ontology driven framework for multimedia information retrieval in a P2P network.

Chapter 3

Content Based Multimedia Retrieval

3.1 Introduction

The recent technological advances have made digital multimedia resources more affordable and easily available for a wide variety of uses. However, accessing the content of these resources is not a trivial task, primarily because of their dynamic and temporal nature. The task of content based multimedia access has recently drawn increased attention in the multimedia information retrieval research community. As previously presented, in this work, we focus on multimedia resources that are generated by scientific conferences and academic events. Nevertheless, the rich information that is implicitly conveyed in these resources could be made accessible by concentrating efforts to create useful semantic annotations and to offer semantic based searching (querying and browsing) and visualization approaches. Several works have addressed the issues related to the Content Based Multimedia Retrieval (CBMR).

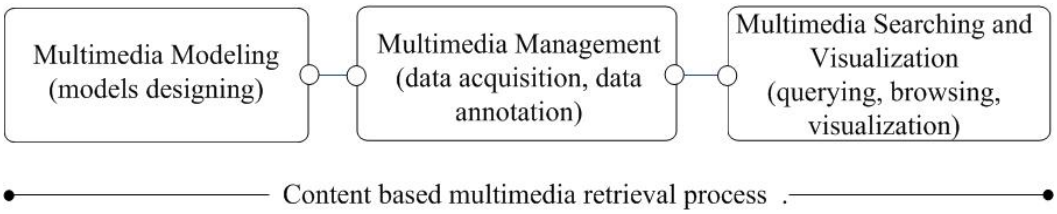


Figure 3.1: Content based multimedia retrieval process

The CBMR process may be represented as the combination of different entities: (1) The multimedia modeling which involves models design, (2) the multimedia management which involves the data acquisition and annotation aspects, and (3) the multimedia searching and visualization which involves the querying, browsing and visualization aspects (Fig. 3.1). The figure 3.2 provides an example of an overview of the CBMR provided by the authors of the Virtual Information and Knowledge Environment Framework (VIKEF) project [VIK]. In this figure we can identify the "information extraction" which is relative to the multimedia management entity, the "knowledge representation" and the "ontology

engineering" which are relative to the multimedia modeling entity and the "smart community services" which are relative to the information searching entity.

In this chapter, we present an overview of current work conducted on content based multimedia retrieval, highlighting three main issues: The multimedia information modeling, the multimedia management (content based annotation) and the multimedia visualization. In the following section, we describe the approaches

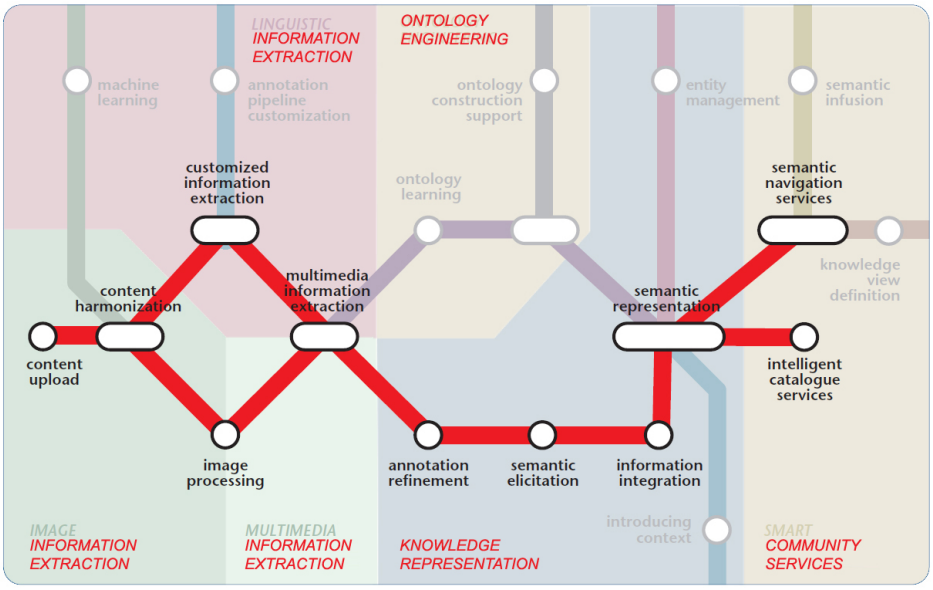


Figure 3.2: Virtual Information and Knowledge Environment Framework (VIKEF) - retrieval process

that are currently used to model scientific and academic information related to an event. In the section 3, we focus on content based annotation approaches to multimedia information retrieval. In the sections 4 and 5, we present an overview of the major content based approach to multimedia searching and visualization approaches. Each section provides motivations for the proposed semantic based approach to multimedia information retrieval .

3.2 Semantic Models

A data model expresses the properties of the data that are to be stored and retrieved. One of the main aims of multimedia modeling is to hide the complexities of the data by using a formal language or semantic schema representation of the data. Multimedia data convey a large amount of information that needs to be organized and structured. A model of multimedia data that is to be used for information retrieval should correspond to users' retrieval requirements. The multimedia models may represent low-level features, such as color, texture and

shape, and high-level features also called semantic features descriptions, such as event descriptions, domain descriptions, topics and relationship. The figure 3.3 presents a layered abstraction of the description that may be associated to a sport video. We can identify in this figure the different level of information abstraction starting from low-level features (bottom of the figure), ending with high-level features referred as customized semantic (top of the figure) [Tjondronegoro 2005b].

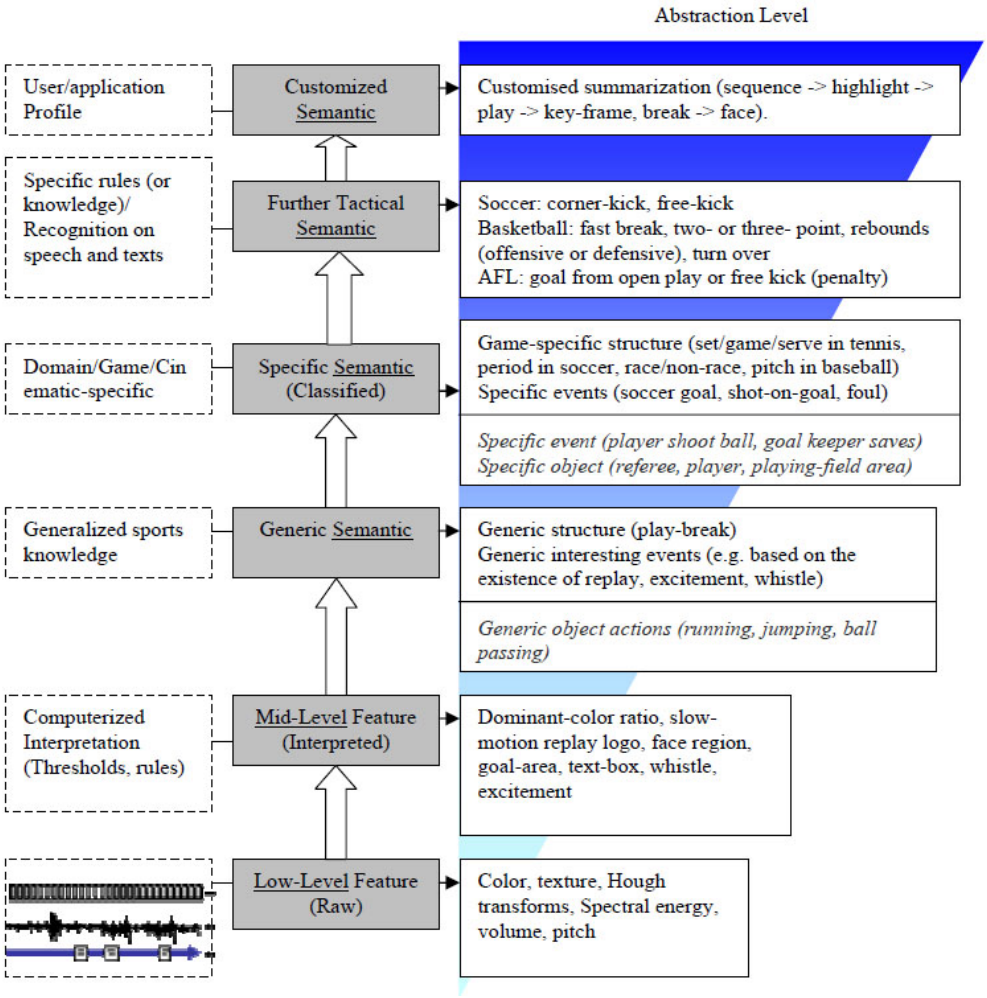


Figure 3.3: Multi-layer abstraction of sport video

Modeling the multimedia semantic content may be more difficult than modeling the low-level features since semantic descriptions may be subjective, thus differ from one person to another. According to [Pridmore-Brown 2010], ontologies explicit specifications of the conceptualizations of domains, they provide a technique in the area of "tacit knowledge" for structuring historically unstructured

and poorly represented data, accessible to both people and applications. To enable computers to access the knowledge conveyed within stored multimedia data, semantic models need to be developed. Additionally an understanding of the information structure shared by both humans and applications may produce new data. With a shared knowledge model, applications will be capable of extracting and aggregating information from different sources, and search engines will be capable of collecting information in a way that provides more useful results for users' queries.

Another advantage of developing semantic models is that domain knowledge can be reused. For instance, many different domains need a representation of the concept of time. This representation includes the notions of time intervals, points in time, relative measures of time, and so on. Once the concept of time is developed in one of the relevant domains, it can be used for other domains as well. Thus the task of building a large ontology (an ontology integrating other ontologies) may be made easier, in fact, several existing ontologies describing portions of a large domain may be integrated into a larger ontology for that domain.

A semantic model allows explicit assumptions about a domain to be expressed, and this makes it possible to modify those assumptions easily as our knowledge about the domain changes [Zhang 2008]. The current state of knowledge representation, as this technology is often called, is comparable to that of hypertext before the advent of the Web: "It is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world" [Berners-Lee 2001]. Controlled vocabularies, taxonomies and models for knowledge management have been developed since the mid 90s [Umar 2009]. Recently standards such as the Resource Description Framework (RDF) [RDF] and the Web Ontology Language (OWL) [OWL] were proposed by the World Wide Web (W3C) [W3C]. Already, many websites, including Wikipedia [WIK] and Linked Data [LIN], share their data through the standard model for data interchange on the web. This model is known as the Resource Description Framework (RDF) and can be queried by both users and automated agents [GAR]. Ontologies are widely used to structure information for purposes ranging from analysis and artificial intelligence to simple categorization, searching, and information retrieval [Antoniou 2004]. As stated by [Kitamura 2003], an ontology, which is a system of fundamental concepts, makes explicit our conceptualization of the target domain and provides us with a solid foundation for building sharable knowledge bases that are more broadly usable than conventional knowledge bases. The following section presents the concept of an ontology.

3.2.1 Ontology Description

According to the W3C specifications, an ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information. "Ontologies include computer-

usable definitions of basic concepts in the domain and the relationships among them" [ONT]. They encode knowledge within a domain and also knowledge that spans domains. The word 'ontology' has been used to describe conceptual systems with differing degrees of structure. These range from simple taxonomies to metadata schemes to logical theories. A significant degree of structure must be provided in ontological models. These models need to specify descriptions for the following kinds of concepts:

- **Classes** (general things) in the domains that are of interest. For instance, "*Wine*" is a class which has two subclasses: "*RedWine*" and "*WhiteWine*".
- **Properties** or attributes and rules that things in the domain may have. For instance, the class "*Wine*" may have the "*Name*" and "*Flavor*" properties.
- **Relationships** may exist among things in the domain. For instance the class "*Wine*" may be related to the class "*WineLocation*" with the relationship "*is produced in*".

There are different approaches to defining the classes, relationships, properties and rules that are needed to represent a domain's implicit knowledge. Indeed every knowledge domain can be described and modeled in different ways. As stated by [Edgington 2004], ontology design is similar to software design in which an iterative approach is used, including a design phase, a development phase, and an iterative feedback and refinement loop. Philip P.-B. presents in [Pridmore-Brown 2010] three different levels of ontologies: (1) Upper level ontologies that describe general or common concepts, (2) middle level ontologies that extend or map the upper level concepts relative to a specific domain, and (3) lower level or domain specific ontologies that define granularities of a domain. This structure offers flexibility and an opportunity to reuse and specialize ontologies defined at the upper level. As the level of generalization of the ontology decreases it becomes more specific and thus its reusability decreases [Kiryakov 2001, T. E. El-Diraby 2005]. In the study described in [Cullot 2003], the authors describe ontologies through three orthogonal criteria:

- **Ontology Focus**: Wordnet [WOR] basically provide definitions of terms. These terms are intended to be used for thesauri. They have a hierarchical structure organized by the relation of subsumption and include other linking relationships for expressing relations such as synonymy. This kind of ontology is usually referred to as taxonomic. This kind of ontology is useful when it comes to sharing information since it provides a reference vocabulary. Other ontologies go beyond simple taxonomies and define properties and relationships between concepts that articulate a rich representation of the target domain.
- **Ontology Range**: The authors of [Cullot 2003] use the term range to refer to the intended use of an ontology. Some ontologies may be designed and used for explanatory purposes such as a service to enable the understanding

of a domain. Such ontologies usually come without associated instances (real world resources described according to an ontology). Ontologies may also be used to support data management services. Such ontologies have associated instances, stored in either a database or a semi-structured data set.

- **Ontology Context**: Ontologies traditionally convey a single conceptualization. However, many different conceptualizations may exist for a specific domain. Each conceptualization defines the view of an user community. Contextual ontologies have been proposed to support alternative views of the world by providing context based definitions. The advantage of a contextual ontology is that it facilitates easy navigation between contexts.

3.2.2 Ontology Languages

Several specifications such as the model's expressiveness should be considered when representing a body of knowledge with an ontological model [Zhang 2008]. Several descriptive languages for ontologies have been specified. The current key languages are the RDF, RDF Schema (RDFS) and the Web Ontology Language (OWL).

- **RDF**: RDF and RDF Schema were the first step towards web based ontology languages. RDF is a data model for describing resources on the Web. An RDF description is a set of triples, where a triple represents the subject, verb and object of a sentence. A triple may also be referred as a set of subject-predicate-object elements (e.g. a conference -subject- takes place -predicate- in a city -object-). Each element of a triple is a Universal Resource Identifier (URI). RDF descriptions can be depicted with RDF data graphs. RDF Schema is an extension of RDF and supports definitions of basic ontology elements such as classes with their hierarchy and, properties with their domain, range and hierarchy. Thus, RDFS can be used to describe taxonomies of classes and properties and is well suited for expressing simple ontologies.

- **OWL**: OWL is a W3C recommendation for a web ontology language and serve as the representation language for OWL-S ontologies. OWL is developed based on the DAML+OIL (DARPA Agent Markup Language+Ontology Inference Layer - DARPA stands for Defense Advanced Research Projects Agency) language which originated by merging two language proposals that aimed at addressing the expressivity limitations of RDFS. OWL enhances the expressivity of RDFS by providing the means to represent relations between classes such as disjointness, union, and intersection restrictions on property values such as cardinality, property characteristics such as transitivity, and more. One major reason for OWL's choice as the representation language for semantic web was that it is compatible with the eXtensible Markup Language (XML) [XML] and the RDF and it provides such additional expressiveness allowing users to formally describe more types of classes, properties, individuals, and relationships than either XML or RDF can. OWL also provides a formal

semantics that gives terms defined in OWL a precise meaning so that they can be used effectively in applications that require interoperability. OWL is a general purpose representation language that provides no special vocabulary for service applications. Thus those who build service applications need to find a service ontology or build their own. OWL provides three increasingly expressive sub-languages: OWL Lite, OWL DL, and OWL Full. OWL Lite provides less expressive power than OWL DL and OWL Full. OWL Full provides maximum expressiveness while remaining computationally completeness and decidable and thus is a good choice for a representation language when efficient support for reasoning is desired.

There are several editing and visualization tools that facilitate ontology development. Some of these provide the means for analyzing, modifying, and maintaining the ontology as it evolves. The main tools in use today are Protege2000 [PRO b], WebODE [WEB] and OntoEdit [Sure 2002]. They integrate enough functionalities to support the major stages of an ontology's life cycle (creation, visualization, and manipulation). Their designs allow the integration of new modules, offering underlying environments that are independent of the description language. Protege was developed by the Medical Informatics at Stanford University. Protege is an open source and standalone integrated software tool used by system developers and domain experts for developing knowledge based systems. WebODE was developed in the Artificial Intelligence Laboratory at the Technical University of Madrid and is also an ontology engineering suite. WebODE functions as a web server with a web interface. OntoEdit was developed at the Karlsruhe University in Germany. OntoEdit is based on a plugin architecture and is an extensible and flexible environment, providing the functionality to browse and edit ontologies.

As stated earlier, the main goal in modeling multimedia information is to represent the complex and rich information conveyed through the digital resources so that the knowledge can be shared. An ontology allows us to define terms for describing and representing a domain of knowledge. In the following section we present the major existing semantic models that describe the knowledge conveyed in scientific conferences, as this domain is the focus of this thesis.

3.2.3 Conference Models

Several works have been carried out to define ontological models for conferences and events, these include the AKT (Advanced Knowledge Technology) Reference Ontology [AKT], the eBiquity Conference Ontology [EBI], and a vocabulary to describe relationships between people [VOC]. According to the detailed study "The ESWC and ISWC Metadata Projects" [Knud Moller 2007], existing models at the time lacked the expressiveness necessary for representing conferences and so more expressive and detailed ontologies were proposed including the International Semantic Web Conference (ISWC) [Knud Moller 2007] and the European Semantic Web Conference (ESWC) [ESW]. The ESWC semantic web technologies project

combined the creation and publication of metadata describing conferences deploying a range of applications (including a semantic wiki, a photo annotation tool, and a semantic search engine) that used the ESWC Conference Ontology and its associated RDF dataset to represent more accurately conferences for delegates. In contrast to other ontologies, the ESWC Conference Ontology explicitly models relationships between people, roles, and events. So for example, the act of giving a paper at a conference is modeled in terms of a person holding the role of presenter at a specific talk event, and there may also be one or more associated artifacts, such as a paper or a slide set. This ontology makes use of the Person and ResearchTopic classes from the Friend Of A Friend (FOAF) [FOA] and Semantic Web for Research Communities (SWRC) [SWR] ontologies, respectively.

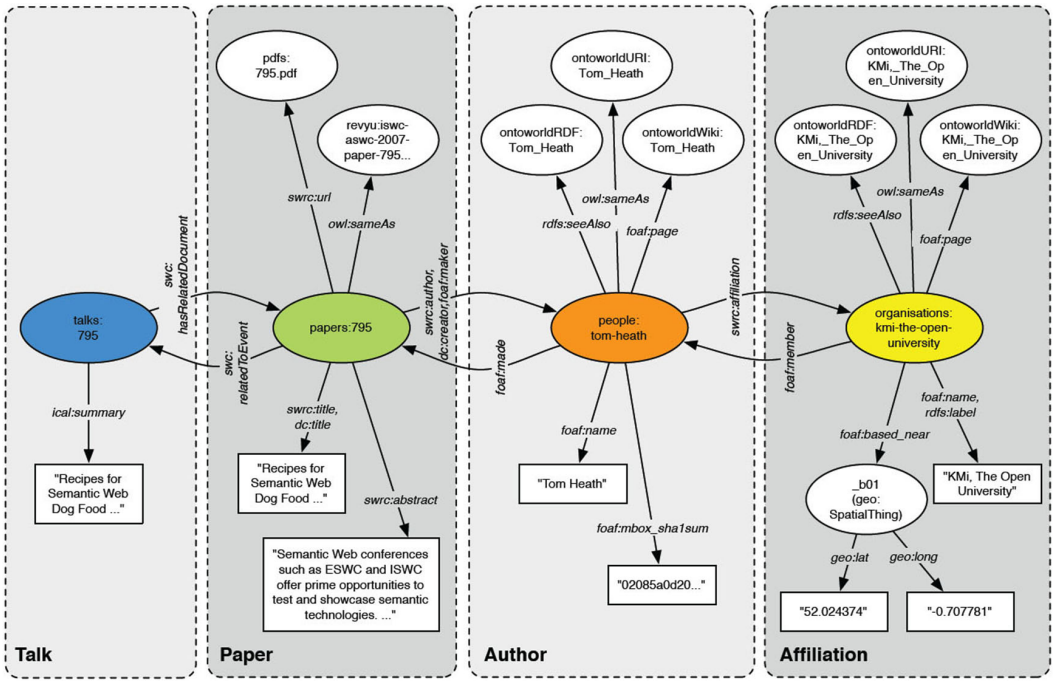


Figure 3.4: International Semantic Web Conference ontology (ISWC)

The ISWC semantic web technologies project based its work on approaches taken and decisions made during the planning of ESWC. Since the ESWC model had already been proved useful and adequate, it was clearly desirable to adopt it for ISWC [Knud Moller 2007]. The main points addressed by the ISWC model were: (1) Increased integration of existing standards and vocabularies, (2) rearrangement of the document concept space, and (3) full use of the potential offered by OWL Lite, such as inverse properties, which aid data visualization. The ISWC model imports the ESWC, adds a number of new concepts and properties, makes some additional statements about existing concepts and properties and deprecates others. The ISWC model incorporates additional established standards for entities such

as the BibTEX schema [BIB] for representing publications. In their work Knud Moller et al. [Knud Moller 2007] propose that academic conferences be represented with three distinct objects: **People**, **Events**, and **Publications**. **People** may take the role of paper authors, committee members, and so on. **Events** may consist of talks (for example paper or poster presentations), conference sessions in which several papers are presented,or various kinds of non-academic events such as meals, social events, coffee breaks also. **Publications** may consist of full papers and poster/demo papers, proceedings, and the like. In addition, Knud Moller et al. defined **Artefacts** such as sets of slides that are not necessarily formally published. The ontology used for the metadata efforts begun with ESWC, was refined during the following conferences, and a relatively stable version has been established [SWC].

The figure 3.4 represents a part of the ISWC conference, it shows the multimedia elements added that can be said about each of the four main entities surrounding a paper which are: In green, the paper that appears in the proceedings - "swrc:inProceedings", in blue, the presentation of the paper - "swc:PaperPresentation", in orange, the people that are related to this paper - "foaf:Person", and in yellow, the information relative to the organization of the event - "foaf:Organization".

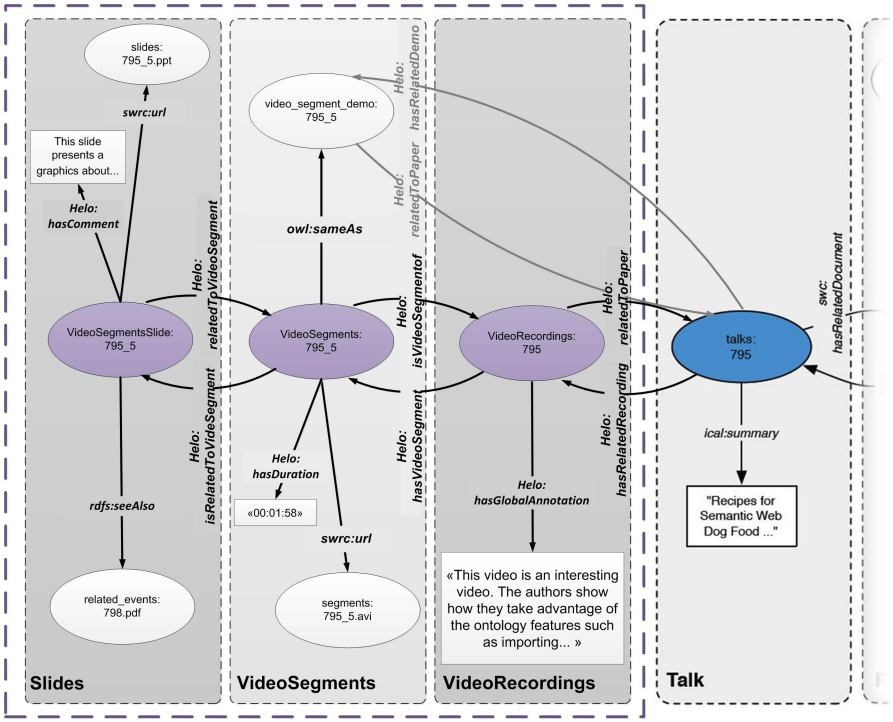


Figure 3.5: Part of the thesis proposed model based on the ISWC ontology

We based this thesis work on the effort done by "The ESWC and ISWC Metadata Projects" since these models have proved their usefulness [Knud Moller 2007]. We used features of the ontology while defining a specific ontological model adapted to the thesis requirements in multimedia information retrieval. The figure 3.5 describes part of the thesis proposed approach which is detailed later in the chapter. It shows the kinds of multimedia elements that have been added in this this work and that can be said about each of the main entities surrounding a paper which are represented in purple: The video recording related to the paper presentation - *Helo:VideoRecordings*, the video segments that are related to the video recordings - *Helo:VideoSegments*, and the slides extracted from the file of the paper presentation - *Helo:VideoSegmentsSlide*.

3.3 Content Based Annotation

In the previous section we presented the main works related to semantic models for conferences. One of the main use of the models is to describe with a common knowledge the content of the digital resources. In this section we present techniques and approaches that have been used to describe the content of multimedia resources.

The content based annotation is an important stage in the enhancement of the multimedia information retrieval process. Indeed, most of the current search engines lack an efficient approach for multimedia retrieval. This is caused by the lack of semantic multimedia content based indexes for such retrieval [Tjondronegoro 2005b]. The semantic information in multimedia data can be expressed either implicitly or explicitly. To understand the implicit semantic information, users need to apply their knowledge although they should be able to understand explicit semantic content more intuitively. Similar to the human perception, the semantic content in multimedia is interpreted as perceived and so there is a need for a common extensible and shared semantic model for describing the semantic content.

A promising and widely spread solution is the content based annotation, which helps the categorization of the multimedia data based on the concepts meanings conveyed therein [Vembu 2006, Dong 2006, Xu 2009, Dimitrova 2002].

3.3.1 Multimedia Annotation

Regardless of the level of annotation (low, mid, or high-level), we may divide annotation into three types: Manual, semi-manual, and automatic. Some studies such as [Xu 2009, Zhang 2002, Maynard 2003], present a slightly different categorization: Manual, rule-based, and machine learning method annotation.

- **Manual annotation** is done by a human whose ability to express semantic information theoretically has the highest accuracy. This approach is a time

consuming task because of the exponential growth of multimedia data being stored for subsequent retrieval. The task may be facilitated with authoring tools such as Semantic Word [Tallis 2003]. Since the use of human annotators can lead to errors (due to factors such as annotator non familiarity with the domain and, personal motivation) and to inconsistency, authoring tools help to integrate common and shared semantic models.

- **Automatic annotation** is performed by machines based on specific algorithms for object recognition or on various techniques such as segmentation and face detection. Automatic techniques are not time consuming for human, but they are mainly applicable to low or mid level feature annotation and therefore are not sufficient for generating the annotations that are needed for semantic content based retrieval [Lawson 2004].

- **Semi-manual annotation** combines the previous methods, employing their respective advantages: The efficiency and the speed of the automatic approach and the accuracy of manual annotations guided by a semantic model. The rule-based annotation and machine learning methods defined by [Xu 2009] are semi-manual approach. The rule-based annotation methods use expert knowledge to classify the annotation. The machine learning methods build large varieties of semantic models and extend these models to the entire multimedia data. The machine-learning methods reduce the necessary manual operations but still rely on learning samples that are provided by humans and annotated manually.

Every existing semantic annotation system requires human intervention at some point in the annotation process [Lawson 2004]. In this thesis we have focused on the manual, semi-manual and automatic annotation methods that are based on semantic models. With the advent of semantic web techniques, ontologies have quickly spread and have been used to design semantic models with concepts, properties and their interrelationships within given domains. They have been developed and used to annotate multimedia over the last decade [Hyvonen 2002, Bao 2004, Dong 2006, Akrivas 2007]. We therefore chose to study multimedia annotation tools that are based on semantic models. In this context the ontology is used to annotate multimedia data in order to incorporate domain knowledge into multimedia indexes, thus increasing the relevance of the information that is retrieved.

3.3.2 Video Annotation

Since this thesis focus is on multimedia resources and specifically on video resources, in this section we present the major works that have been conducted on video annotation. The video annotation is crucial to the success of the video search and retrieval. Precise video annotation and semantic reasoning over these annotations will produce, more efficient and accurate search results. Video annotation is the process of making time based notes corresponding to elements of the video

footage, usually a description of the content for subsequent retrieval [Correia 1999]. Schroeter et al. [Schroeter 2003] describe the video annotation as attachment of subjective comments, notes, and explanations to a document or a selected part of a document without actually modifying the document.

Several projects have been directed towards handling multimedia annotation, these include the K-Space Annotation Tool [Wilkins 2007] developed at the University of Koblenz-Landau, the VideoAnnEx IBM MPEG7 Annotation Tool [VID a], the Caliph & Emir [Lux 2003] project from the Graz University of Technology, the Muvino [Boszormenyi L. 2005, MUV], the iFinder by Fraunhofer [IFI], JOANNEUM Research [JOA], Educational LANguage (ELAN) [ELA], and the EXtensible MARkup Language for Discourse Annotation (EXMARaLDA) [EXM]. Many tools are dedicated to video annotation :

- M-OntoMat-Annotizer [Petridis 2005]: A user-friendly tool developed within the aceMedia project that allows semantic annotation of images and videos for multimedia analysis and retrieval. It is an extension of the CREAting Metadata for the semantic web (CREAM) framework [Handschuh 2003] and its reference implementation, the OntoMat-Annotizer. The Visual Descriptor Extraction Tool (VDE) was developed as a plug-in to OntoMat-Annotizer and is the core component for extending its capabilities and supporting the initialization and linking of RDF(S) domain ontologies with low-level MPEG7 [organisation for standardisation 2004] visual descriptors.

- Advene [ADV]: Annotate Digital Video, Exchange on the NEt, a tool that provides a model and a format for sharing annotations about digital video documents (movies, courses, conferences, and the like), along with tools for editing and visualizing the videos combining both the annotation and the audiovisual document.

- Anvil [ANV]: A free video annotation tool that is used at research institutes world-wide (Fig. 3.6). It offers frame-accurate, hierarchical multi-layered annotation that is driven by user-defined annotation schemes. An intuitive annotation board shows color-coded elements on multiple tracks in time-alignment. Special features include cross-level links, non-temporal objects and a project tool for managing multiple annotations. Originally developed for Gesture Research, Anvil has also proved suitable for research in human computer interaction, linguistics, computer animation and many other fields.

- Annomation [ANN , Lambert 2010]: A tool that handles the users inputs annotations. Annomation provides a simple web browser interface split into four main areas (Fig. 3.7): A video player (top-left), a list of videos, and existing annotations for the current video (top-right), a set of controls for the video player, and input widgets to enter the annotations (across the centre), and a set of panels to aid in finnding suitable linked data annotations (at bottom).

- Vannotea [Schroeter 2003]: A tool that enables real time collaborative annotation of video content. Vannotea architecture is based on three servers. The indexing, annotation and application servers. The indexing server is used for indexing and video content retrieval. The annotation server is used to store,share and retrieve the annotations. The metadata are based on MPEG7 standard.

- PAMS [Su 2010a]: Personalized Annotation Management System, a system consisting of two parts. An annotator side and a server side. On the annotator side, handlers collaborate to process annotation creation, retrieval, discussion, and management. Annotators can create, edit, and retrieve their own annotations. This annotation management model has three types of metadata (annotation metadata, annotated object metadata, and annotated document metadata) derived from Dublin Core for annotation management. The server side supports an anchoring position mechanism, a document manager, an association manager, and a user manager.

- POLYSEMA MPEG7 Video Annotator [POL]: A tool that performs MPEG7 based annotations for video content. This tool supports high-level multimedia annotation, based on Part 5 (Multimedia Description Scheme, MDS) of the MPEG7 standard. High-level annotation refers to metadata that describe high-level concepts rather than structural, visual or audio features. This tool currently takes as input an existing MPEG7 document that is assumed to contain basic information about the video segments (start time and duration), possibly obtained from a shot segmentation tool or algorithm. The tool supports annotation both for the entire video and for discrete video segments using RDFS/OWL domain ontologies as controlled vocabularies.

In this thesis work we are specifically interested in conference video recording annotation based on semantic techniques. For this we needed a modular tool that supports the integration of new ones such as specific module for conference video recording segmentation. We needed also a tool that support semantic annotation techniques such as RDF and OWL. As POLYSEMA answered these requirements, and since it is an open source tool with a flexible architecture, we choose in this thesis work to extend it. The details are presented in the chapter 10 section 10.4.1.4.

3.4 Content Based Browsing and Visualization

Because of the substantial challenges in bridging the "semantic gap", content based multimedia searching (querying and browsing) has been a major focus of researchers since the 90s. The gap between low-level multimedia features and high-level semantics needed to be bridged in order to support intuitive and user oriented searching [Li 1995]. However, despite promising advances, effective multimedia search remains an open issue. Until recently, text based searching was the most highly

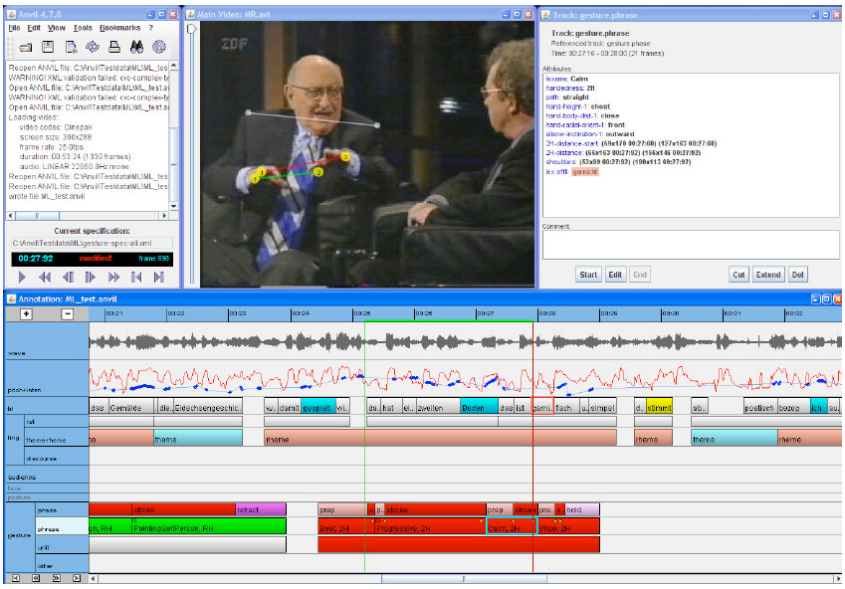


Figure 3.6: Anvil - video annotation tool

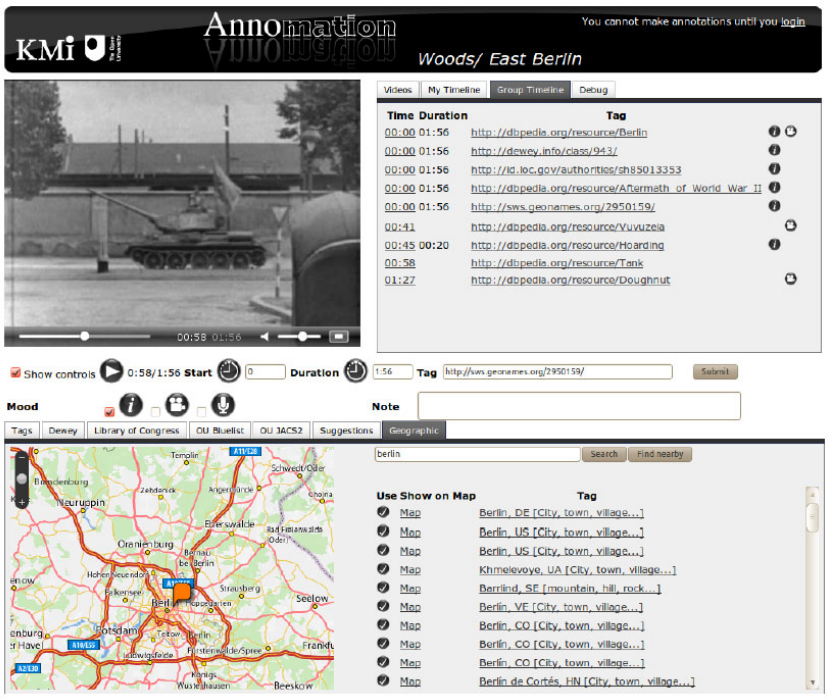


Figure 3.7: Annomation - video annotation tool

developed area and a few other types of search were supported for specific multimedia collections [Tjondronegoro 2008]. As noted by [Tjondronegoro 2009], web search can be vastly improved by focusing on users' search behavior and content based information retrieval. In fact, future information retrieval studies will need to understand and take into account users' search behavior, in order to narrow the "semantic gap", and they will need to consider the integration of different media [Tjondronegoro 2009, Kherfi 2004]. The retrieval process can be further expedited by addressing the issue of content based visualization. Visual representations can be used to aid the analysis of visualized information and to facilitate the discovery of relationships, clusters and gaps. In this section we focus on the content based browsing and visualization approaches.

3.4.1 Faceted Search

Faceted search is a user interface paradigm based on flexible classification [Polowinski 2009, Schmidt 2010]. Hearst describes facets as a "a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain" [Hearst 2006], while LaBarre characterizes facets as representing "the categories, properties, attributes, characteristics, relations, functions or concepts that are central to the set of documents or entities being organized and which are of particular interest to the user group" [Barre 2006].

Web search engines are one of the most widely used technologies of our new digital era. As reported by Fallows et al. in their study (based on the American population) [Tjondronegoro 2005a], already in 2005, over 80% of the computer users have used a search engine to seek information online. Everyday, more than 60-million users query web search engines. Until recently, most of the web search research focused on general web search, with less attention paid to multimedia search [Tjondronegoro 2009]. Tjondronegoro et al. surveyed the multimedia search functionalities of search engines such as Google, Yahoo and some other and showed that fewer than 5% of the investigated retrieval systems provided content-based search while a few search engines offered multimedia search with generally limited functionalities [Tjondronegoro 2009]. The ability to search multimedia data requires meaningful query formulations. The latest studies, such as the "Usability studies of faceted browsing" show that faceted search, also known as faceted browsing offers an efficient approach for formulating complex and meaningful queries [FAG].

Faceted search is used to explore structured resources and iteratively perform and refine complex queries (e.g. Fig. 3.8). At the outset a complete set of items is presented, and this set reduced to subsets as the query is refined by the faceted search. This gives the users a flexible and personalized system of navigation. The faceted search offers relevant subcategories, allowing users to see an immediate visual overview of search results and to refine or expand their search. Faceted data browsing was introduced in the Flamenco Search Interface project

[FLA a] which was designed to help users navigate through large quantities of data in a flexible manner and without getting lost. In the literature review [FAG], the authors state that numerous studies in the information science literature report positive results for faceted browsing and build a solid case for using facets in search interfaces: "Facets are useful for creating navigation structures", "Users are faster when using a faceted system", "Faceted categorization greatly facilitates efficient retrieval in database searching". However, to be effective, the organization of a hierarchy must match or closely match the users' mental model [Perugini 2010].

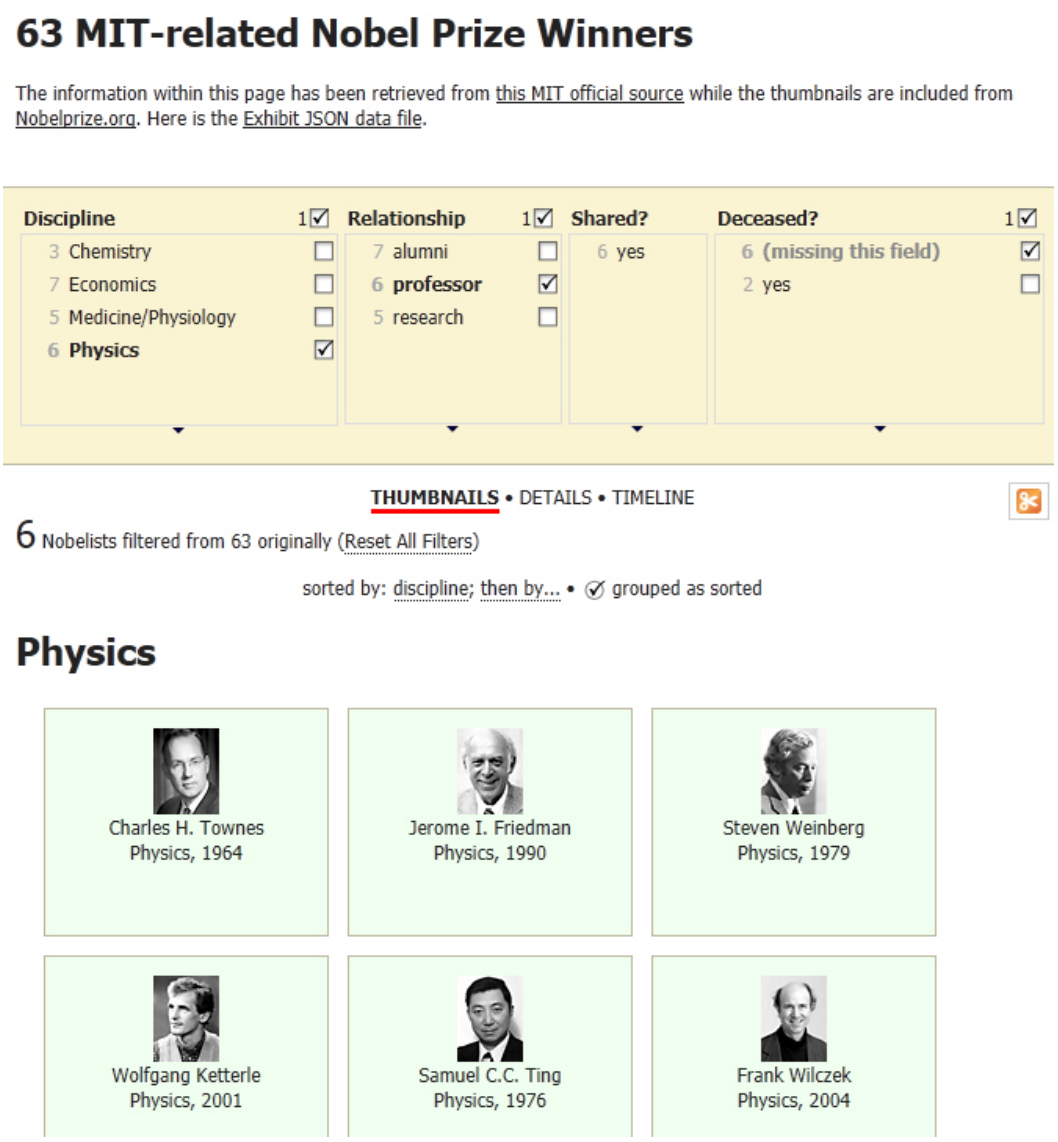


Figure 3.8: Simile - example of faceted search

In their study [Perugini 2010] Perugini S. et al. provide an accessible overview of faceted browsing and search and raises a set of questions that should be explored. The research questions involve various aspects of faceted search such as system goals, targeted users, and empirical results from Human Computer Interface (HCI) studies: Research in faceted browsing and search currently focuses on improving user interfaces and connecting faceted search not only to other search techniques, but also to other areas of computer science such as data mining and clustering.

Some broad directions for future research have been proposed including:

- User interface and visualization enhancements to augment standard faceted browsing and search interfaces: These include improving visualizations for faceted filtering, restrictions [Tvarozek 2008, Weiland 2008] and automatically constructing faceted interfaces [Dakka 2005].
- Knowledge discovery in faceted browsing and search: Some projects deal with augmenting faceted interfaces with visualization and interactive facilities for knowledge discovery [Lee 2009].
- Integration with alternate search paradigms: Some research addresses how to best combine faceted search with other search paradigms such as full-text search (e.g., Google) and content-based query-by-example search [Tvarozek 2008, Wilson 2008].
- Application domains beyond websites and digital libraries: Integrating faceted techniques into file systems is a promising approach to enabling users to familiarize themselves with the entire contents of large disks [Fatemi 2003, Koren 2007, Huynh 2007].

In this thesis work we use the faceted browsing approach, whose main advantage is that it can accommodate different users who may approach the same information from different perspectives. This is an essential feature for successful information retrieval [GLA]. As we focus in this thesis work on multimedia information in the following section we present the main research works and techniques that have been proposed for multimedia visualization.

3.4.2 Multimedia Visualization

The terms ‘data’, ‘information’ and ‘knowledge’ are often used in discussions of visualization issues. As stated in [Chen 2009], in many cases each term refers to a level of abstraction. In computing, ‘data‘ and ‘information‘ are often used to indicate the same level of abstraction with respect to visualization. Data or information visualizations tend to use rather simple but clearly defined visual languages, while knowledge visualizations however use more abstract and flexible languages [Burkhard 2005].

Visual representations are used to guide the analysis of the visualized information and to facilitate the discovery of relationships, clusters, gaps, and additional information. The advent of graphic techniques and tools has encouraged researchers to develop novel approaches and techniques for visualizing and interacting with large amounts of data. The visualization approaches are mainly focused on visual abstraction and metaphors and their corresponding interactive models, with the aim of enhancing the exploration and analysis of visual information. Information visualization is distinguished by a clear focus on information retrieval [Sabol 2008]. Most information stored today has a multimedia format and most of the contents are unstructured data that can not be accessed in a unified manner [Sabol 2008]. Visualizations make use of the human cognitive processing system in order to create and convey content more efficiently. Information and knowledge visualizations employ similar techniques. Using mapping rules, resource objects are translated into visual objects with meaningful representations, offering easy and comprehensive access to the target subject matter [Hirsch 2009]. The figure 3.9 presents a set of visual wiki applications.

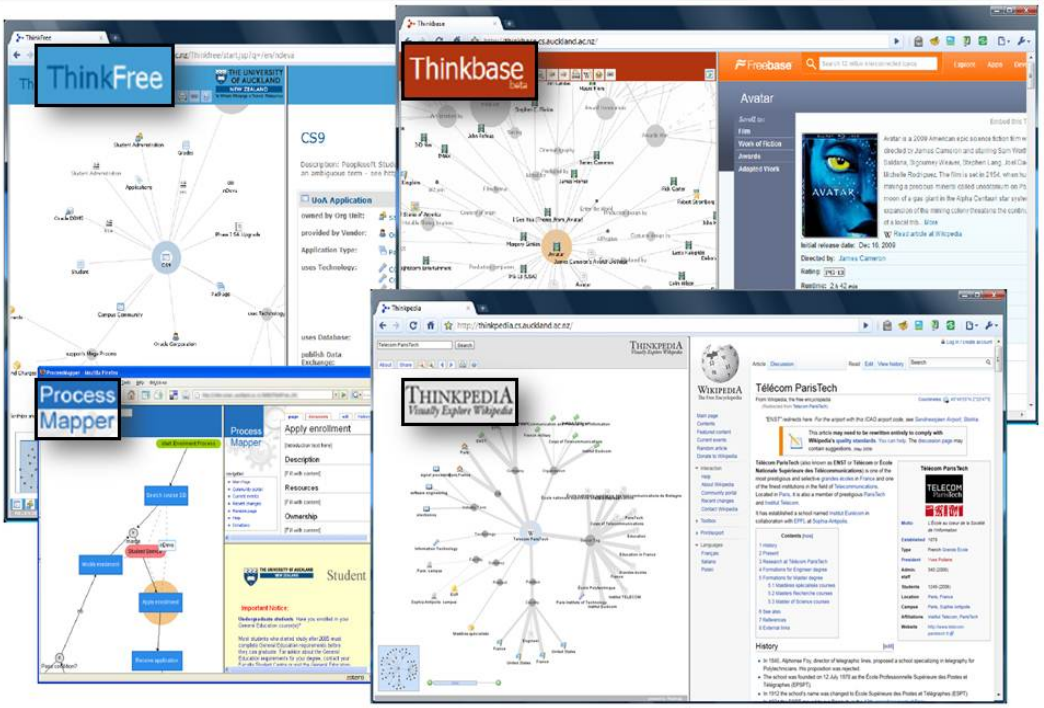


Figure 3.9: Visual wiki applications - Thinkbase, Thinkpedia, ProcessMapper and ThinkFree

An increasing amount of multimedia content is undergoing annotation using semantic techniques and tools. In addition to using multimedia indexing and

annotation techniques, multimedia visualization approach should be aimed at helping users access and understand retrieved multimedia information [Sabol 2008]. When users get in front of a search engine interface they rarely have a clear idea of how to proceed to achieve their goals, and for this reason the user interface should help them express their queries, retrieve the information for which they are searching for and keep track of the progress of their search [Baeza-Yates 1999].

A good interface with information visualization will not only make a large amount of relevant information visible, but will also convey more information than raw data. In "Knowledge Mining and Visualization on NewsWebpages and Large-Scale News Video Database" [Luo 2008a], Hangzai Luo et al. describe several categories of research. One of these categories is geographic information extraction and its representation in a world or regional map. Map representations can be used not only for various applications but also for browsing and navigating. Other types of representations include timeline based representation such as ThemeView, Storyboard representation and hybrid visualization. In his research, Daniel Keim [Keim 2002] classified visualization techniques as in the figure 3.10. Multimedia

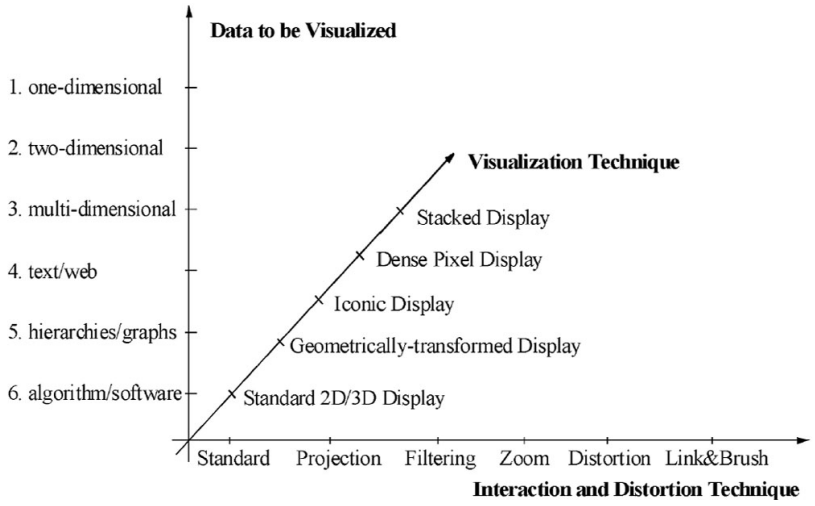


Figure 3.10: Classification of the visualization techniques according to the study of Daniel K.

data, and video resources in particular, present unique challenges due to the dynamic nature of the media and the lack of advanced and precise methods for handling non-textual features [Bagdanov 2007]. Several approaches to multimedia visualization have been proposed including Haystack [Quan 2003], mspace [m.c. schraefel 2005, Luo 2008b], in which RDF data are represented as a collection of facets. Other approaches use graph based representation [Sundara 2010], namely RDF Gravity [GRA], IsaVis [ISA], GViz [Visualization 2005] and Ask-GraphView [Abello 2006]. Since graph based visualization may present the drawback of visual result obstruction or application performance when displaying, some works have

proposed combining these two approaches.

The authors of [Luo 2008b], propose the 3**S** approach (Subsets, Summaries, Sampling):

- Subsets: The ability to show only subsets of a graph. The subsets could be either static subsets such as relationships among instances of a class, or dynamic subsets such as the result of a SPARQL (SPARQL Protocol and RDF Query Language) query [SPA].
- Summaries: The ability to replace portions of a graphs with their respective summaries, which can subsequently be incrementally expanded on demand. Such a graph, referred to as a hybrid (detail-summary) graph, provides a compact representation of the data.
- Sampling: The ability to present a small representative subset of a large collection of instances, for which sampling may be used. A sample of the graph may be visualized using various node-based and edge-based sampling methods.

Depending on the type of information to be visualized, different visualization techniques may be used. An interesting approach that combines faceted search and graph techniques has been implemented in some projects. The following section presents current approaches to visualizing information in the context of scientific and academic events.

3.4.3 Conference Information Browsing and Visualization

There are several applications in the field of scientific conferences that exploit knowledge about the context. In this section we present the projects that we believe are the most relevant to this thesis work: DBLPVis [VIS], ARNETMiner [ARN], RKBexplorer [Glaser 2007] and the Videolectures [VID b] portal:

- DBLPVis (Fig. 3.11) is a tool based on data collected in the DBLP (Digital Bibliography and Library Project) [DBL] project, at the University of Trier. Its main purpose is to support research of a large amount of information by linking conferences, people and words. It can display graphs presenting the results. For example, when a user searches for a person, DBLPVis constructs a graph presenting the relationships between the retrieved person and others to whom the person is related.
- ARNETMiner (Fig. 3.12) is specific to the social networks research domain. It allows us to visualize a social graph that illustrates links between people along with the nature of these links. The relationships can only be displayed one type at a time (advisor, advisee or co-author), and only the relationships between the focused person (the person for whom the search has been made) and people directly connected with that person can be viewed.

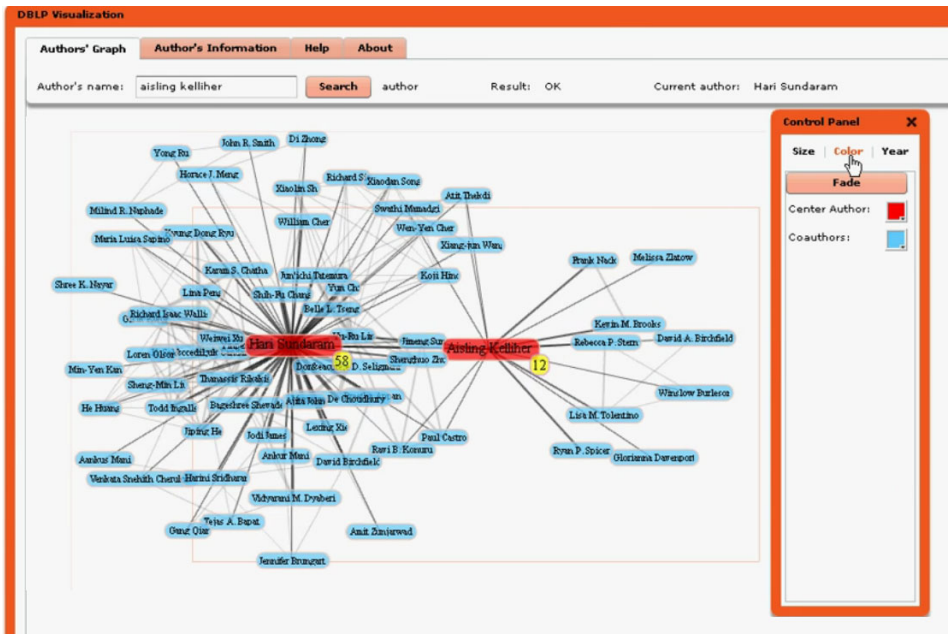


Figure 3.11: DBLPViz interface

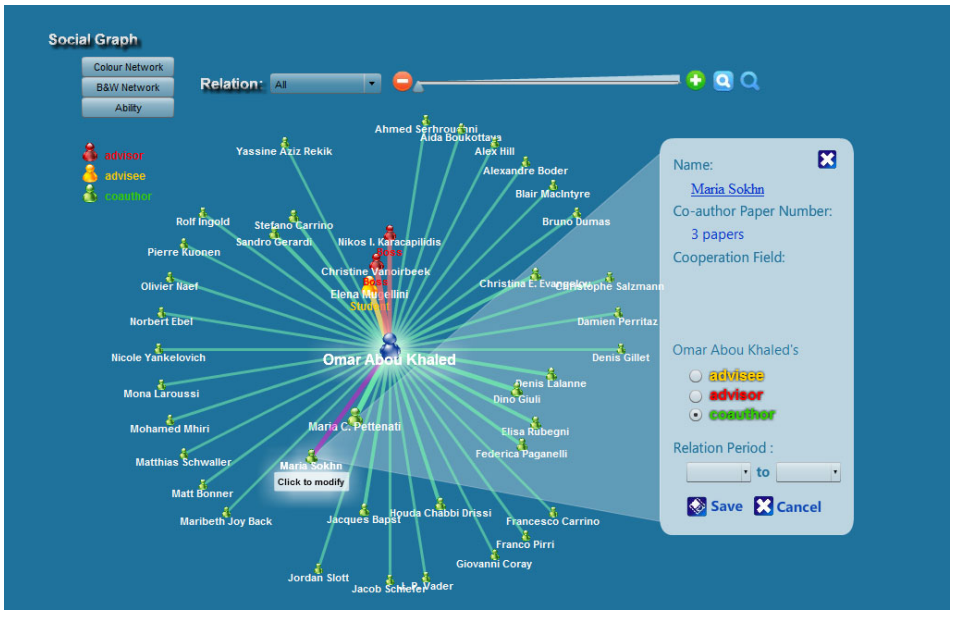


Figure 3.12: Arnetminer interface

- RKBexplorer (Fig. 3.13) is a system for publishing linked data to semantic web standards, providing a browser that allows users to explore this interlinked web of data, primarily in the scientific domain. The RKBExplorer browser presents a unified view of different repositories such as ePrint along with related data from other sources such as DBLP and DBpedia (Data Base wikipedia - a semantic web version of Wikipedia).

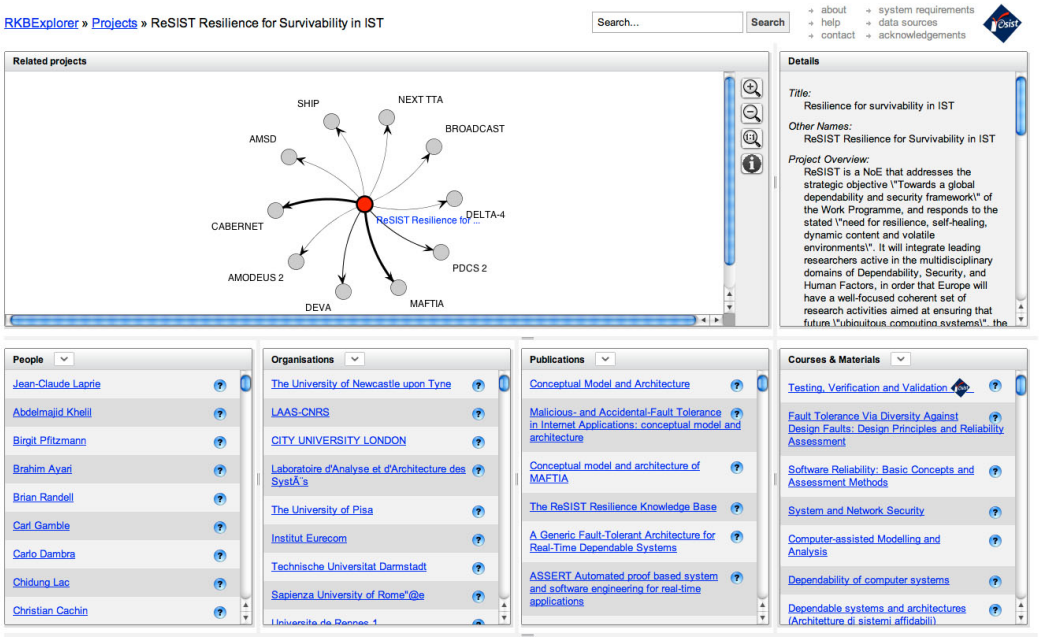


Figure 3.13: RKBexplorer interface

- Videolectures (Fig. 3.14) is an open access educational video lectures repository. The portal is aimed at sharing by providing didactic contents to the community.

Several works such as Microsoft Office SharePoint Server (MOSS) [MOS], EBay [EBA], Amazon [AMA], etc. presented faceted approach. We focused on those that present higher similarities with this thesis. DBLPViz, ARNETMiner and RKBExplorer present faceted approaches to browsing that focus on the scientific community (persons, relationships between them, and so on). However, none of these is based on a conference semantic model that can describe multimedia information such as video recordings and, slide presentation. RKBExplorer has an interesting approach to integrating information from different resources such as papers and projects but it does not search recorded talks (or related data). This framework for information retrieval does not take advantage of the multimedia information (more specifically, information contained in video recordings) that is generated during the life cycle of a conference. On the other hand, Videolectures integrates a large number of interesting video conferences. It offers a faceted search with respect to topics,

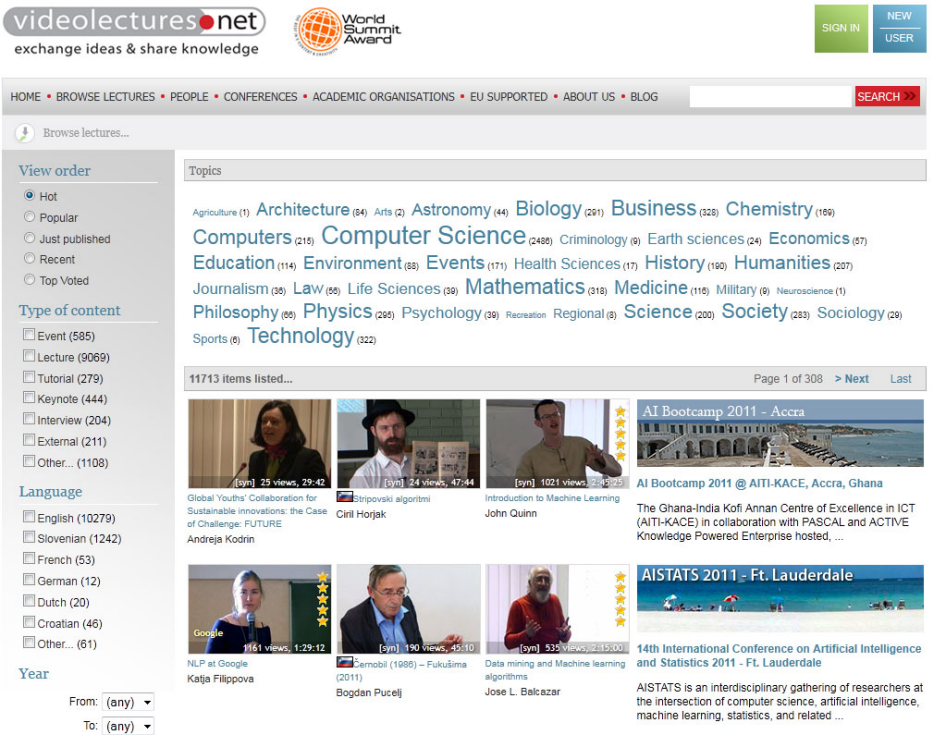


Figure 3.14: Videolectures portal

participants, event types and time periods. However, Videolectures does not utilize a semantic model for conferences and consequently is not able, either to dynamically present the relationships between the various resources (participants, talks, events, etc.) or to take full advantage of the information generated during a conference life cycle.

3.5 Content Based Querying

Having presented the content based browsing and visualization, in this section we turn to the content based querying, an issue we believe interesting to be presented in the context of the information multimedia retrieval state of the art. However, due to the accorded time this issue has not been studied in detail along this thesis work.

In their study [Marchand-Maillet 2000], the authors group queries into different categories:

- **Visual query**: Content based retrieval systems support the searches for visual objects. One logical approach is to search based on examples of the desired document. There are several such research projects and commercial products. In the simplest case the user is asked to enter a single similar image while

in other cases the user can upload more than one image. These systems adjust their algorithms based on users' behavior during retrieval process. Some of them also offer users the opportunity to sketch their image. The major drawback of these tools is the complexity involved in the query formulation [Marchand-Maillet 2000, Tjondronegoro 2009].

- **Motion query**: Motion-based query may be an interesting innovation for video search engines, however, formulating such a query is not trivial. The user is asked to describe motions, or to choose global motion orientation within each quadrant of the motion frames [La Cascia 1996], or to represent a question with a sketch drawing [Chang 1998]. The major drawback of this approach is the difficulty in designing a graphical user friendly interface.

- **Textual query**: Textual queries may seem easier for users of search engines. The user can use keywords to express high-level concepts which would be difficult to express through visual or motion queries. Textual annotations should be related to high-level semantic concepts. However, high-level concepts are in principle difficult to be automatically created. Thus, it is necessary that the user and system collaborate in performing this task.

- **Combining query types**: In the study [Marchand-Maillet 2000], the authors proposed using combined query systems that take users' requirements into account as much as possible. This approach may be made possible with new technologies and has recently been reconsidered in the area known as multimodal querying [Rasiwasia 2010].

Other studies such as [Fatemi 2003, Najood Al-Ghamdi 2010, Saleh], focus on the semantic query approach. In multimedia retrieval systems, users need to query at a high semantic level rather than use low-level features such as color and shape to describe the content. A set of semantic query languages have been proposed to address this need. These query languages extend traditional database querying languages by supporting the formulation of queries that are based on the structure of a multimedia document, its spatio-temporal relations, and so on. Several projects use the semantic web technologies approach for querying web data sources [Saleh , Weiand 2008]. This approach aims at advancing current research by taking advantage of an additional layer of machine interpretable metadata [Najood Al-Ghamdi 2010].

3.5.1 Querying Heterogeneous Data

Integrating data from multiple heterogeneous sources leads to the coexistence of different data models and scheme and therefore different query languages. This situation raises the issue of heterogeneous data integration, management and retrieval. To deal with this issue, classical approaches exists such as **wrapper-mediator architectures** where the wrappers are responsible to adapt local data such that they adhere to the common data model once exported and the mediator stores a global integrated schema and its relationships to the local scheme

exported by the wrappers (Fig.3.15) [Wiederhold 1992, Izza 2006], and recently approaches based on **peer data management** where no common global schema is required and dataspaces where the emphasis goes to the data co-existence providing base functionality over all data sources, regardless of how they are integrated [Klette 1991, Mauroux 2006, Aberer 2011] have been proposed.

The structural and semantic heterogeneity of data makes the development of custom solutions for querying these data time-consuming and complex. This issue can be divided into three main parts: Dispatching relevant search information depending on the data sources, creating database-specific queries and merging results from several sources. Several works have addressed these issue. Munir et al. [Munir 2007] and Tzitzikas et al. [Tzitzikas 2005] both propose architectures with global merged ontologies/taxonomies. Each source of information has its own local ontology/taxonomy. Domenig R. et al. describe the SINGAPORE architecture for which an administrator manually defines similarities and conflicts between the schemes and rules used by different sources [Domenig 2000] . Sattler et al. use the classifications Concepts, Properties and Categories for mapping [Sattler 2005]. As a new source is registered, a map is created between the source and the supported classification. In XLive work (Fig. 3.16) [tram Dang-ngoc 2005], an XML Light Integration Virtual Engine, the structures of all connected sources are simply recorded in an XML configuration file. Hammer et al. [Chawathe 1994] present the TSIMMIS project which uses a schema-less approach that is well suited for sources with dynamic contents. It does not use a global schema describing the sources or a fixed schema for the data.

3.5.2 Querying Systems

All the heterogeneous data source integration systems share a common approach which we refer to as the *fixed point*. By this we mean that a part of the data is explicitly filled in by users. There is either a mapping to be created (a mapping of terms between local and global schemes, taxonomies, ontologies and other concepts), a global schema to be built or a source structure to be explicitly described:

- SemanticLIFE: The SemanticLIFE (Fig. 3.17, 3.18) [Hoang 2006] project at Vienna University of Technology [VIE] attempts to achieve to the vision of Memex of Vanevar Bush [MEM]. A virtual query system and a virtual query language have been developed as a part of this project. The system is fed with data, and external sources can also be queried. It is based on the ontology to remove ambiguities occurring in user queries.

- SIRUP: The semantic integration of heterogeneous data is also the goal of the SIRUP [SIR] project lead by the University of Zurich [UZH]. A semantic multi-data source language is used to declaratively manipulate the so-called

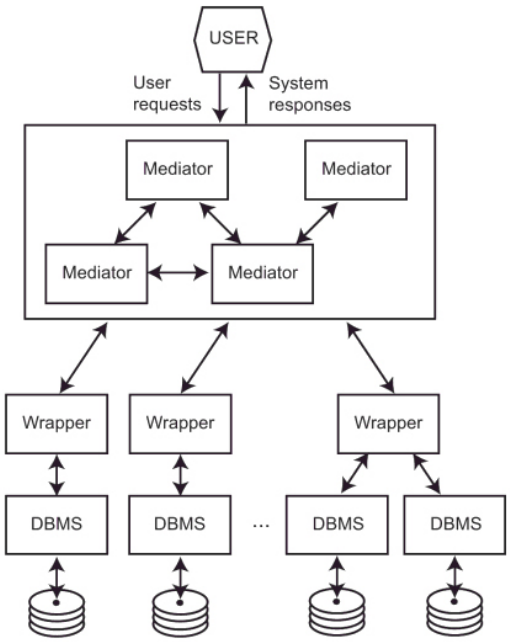


Figure 3.15: Wrapper-Mediator architecture

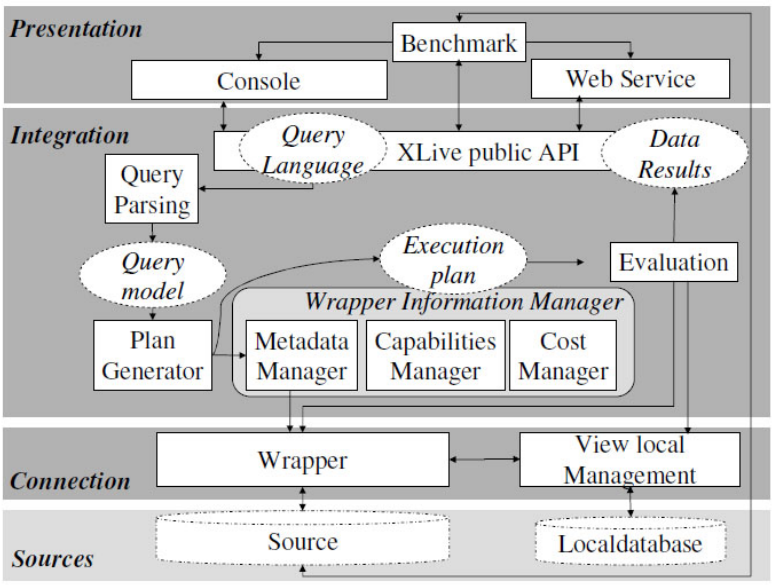


Figure 3.16: XLive architecture overview

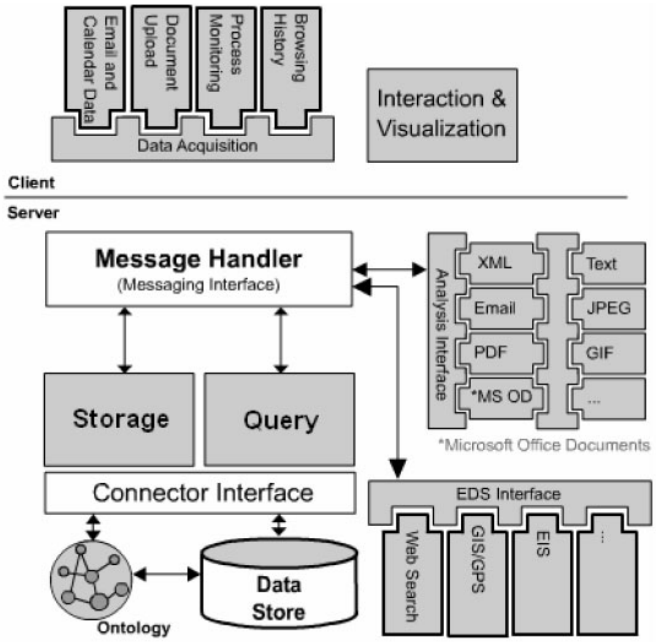


Figure 3.17: SemanticLIFE framework overview

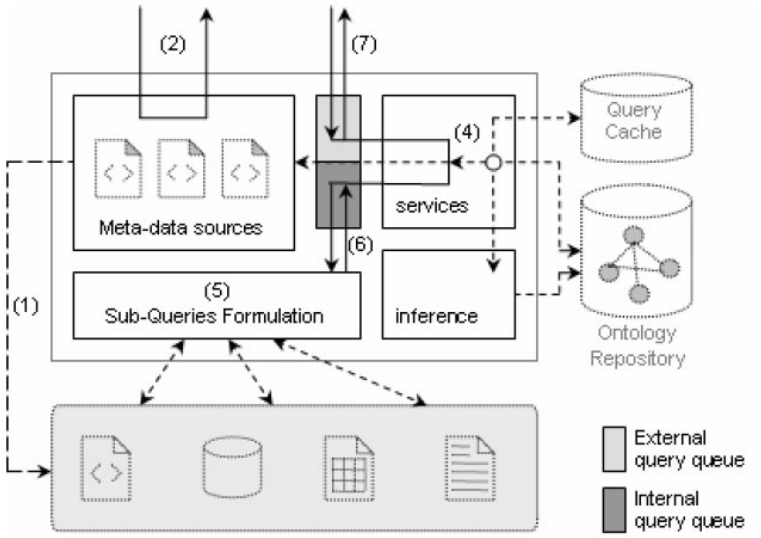


Figure 3.18: SemanticLIFE virtual query workflow

"Concepts". This project provides explicit, query-capable semantics by connecting "Concepts" to the concepts of ontologies.

- NEPOMUK: The NEPOMUK framework [Sauermann 2005] integrates the Gnowsis Semantic Desktop [Sauermann 2003] which aims to add web semantics to classical desktop applications. The data are integrated into a desktop web server using adapters, enabling the user to navigate through documents.
- Haystack: This MIT project [MIT] has investigated new approaches for users to manage information using relationships. In this project, they have developed several applications within the Haystack Project [HAY].

Each of the described systems has a specific approach that falls under two main categories: Either the data are directly fed into the system or the system accesses the different existing sources to retrieve data. Hybrid solution can be found as well. In this thesis, we propose, based on the approaches described above, an autonomous (data structure independent) query engine able to query over heterogeneous data sources by taking advantage of semantic and clustering techniques.

3.6 Conclusion

The oft-cited rapid explosion of information has been accompanied by a diversification of the type of information. Multimedia data are rapidly becoming the predominant form of shared information. Multimedia data sets are characterized by their heterogeneous nature and complex structure. Based on the state of the art presented above we can identify three main issues to be addressed with respect to content based multimedia retrieval. (1) The content based management, (2) content based querying and (3) content based visualization.

Without adequate techniques and tools that enable easy access to the content of multimedia data, the growing information is hardly usable. Moreover, low-level features that are mainly extracted by automatic techniques, are not sufficiently rich to describe multimedia data at a high-level of semantics required by end users. High-level semantic descriptions are mainly provided by manual efforts. Semantic annotation techniques and tools have been developed to reduce the burden of the manual tasks. A few projects try to combine automatic and manual approaches, to the best of our knowledge none of the existing projects made use of the existing information conveyed along the life cycle of a scientific conference.

At the visualization level, several techniques attempt to provide insight into large multimedia data sets. The information is represented to the user using the most suitable and intuitive approach. As stated by [Sabol 2008], the human factor plays an important role, including the ability to navigate, to relate, to remember and to understand complex information, and possibly to collaboratively share and utilize knowledge. Thus advanced applications in the multimedia visualization area

have a significant potential for improving tasks related to knowledge management by making multimedia information more accessible, understandable, sharable, and dynamic.

Studying users' interactions with search engines may potentially impact the development of more effective multimedia search systems. Moreover, semantics techniques appear to be driving multimedia information management and retrieval to its next phase. Thus, challenges in developing semantic techniques have arisen in many areas such as data modeling, query processing, knowledge representation, and the like.

In order to take advantage of existing techniques and approaches to the different component of the information retrieval process (information management, information searching and information visualization), we propose an integrated framework driven by an ontological model. The following chapter presents an overview of the most relevant frameworks in the scientific and academic contexts.

Chapter 4

Multimedia Information Retrieval Platforms

4.1 Introduction

In the previous chapter we cited several projects aimed at improving content based multimedia retrieval. Advanced specialized content based video retrieval techniques [Wang 2011, Su 2010b] have been developed. Several studies proposed integrated platforms/frameworks that take advanced of these techniques in order to provide efficient information retrieval. As stated earlier, the characteristics of conferences make the use of the potential of the Peer to Peer (P2P) promising for the enhancement of the multimedia information retrieval. Indeed, in the conference realm, the resources are distributed. This recall consequently to the necessity of sharing the distributed information and knowledge within the community. The large scale of the scientific community and its computational resources can be exploited by using the P2P potential, hence improving the multimedia retrieval process [Ramzan 2010]. In their work [Hao Ding], Hao D. et al. aimed at comparing and identifying the strengths and weaknesses of both P2P and semantic web technology, in order to show the advantage gained by combining semantic search in a large scale distributed environment (Fig. 4.1). In our context, the principle of collaboration and sharing may be of interest [Klampanos 2003]. Existing P2P already form very large networks with hundreds of thousands or even millions of computers participating. P2P networks can act as a fundamental platform for the searching and sharing of distributed information by using the semantic web technology [Hao Ding]. By combining the resources of a very large number of peers, we can expect a qualitative shift in the information retrieval systems of the future.

As the aim of this thesis is to enhance the retrieval of scientific information, we have focused this thesis study on platforms developed for information retrieval in the domains of scientific and academic events. In the second section of this chapter we describe three frameworks that we consider to be the most relevant to this thesis work. In the third section we introduce the P2P approach to information retrieval.

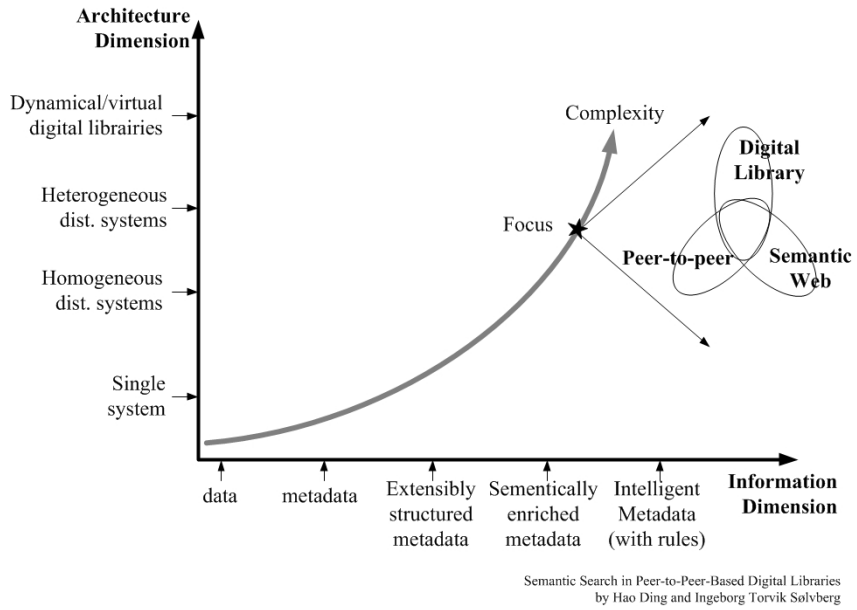


Figure 4.1: Combining P2P and semantic web for constructing digital libraries

4.2 Frameworks

Several frameworks have been designed for multimedia information retrieval. These include [VIK , MAT , Nan 2009, Carbonaro 2008, Z. 2008, H. 2006, PHA , Fatemi 2003] to cite a few. Some of these frameworks, such as PHAROS [PHA], are dedicated to non-domain specific multimedia information retrieval. Tjondronegoro [Tjondronegoro 2009] discusses two different types of framework: (1) The features based approach and (2) the domain-based approaches. He demonstrates the comparative efficiency of the latter type of framework approach. Indeed, a domain specific approach relies on the knowledge of the target domain to enhance high-level information extraction. As noted by [Tjondronegoro 2009], the main idea behind the domain specific approach is to use domain knowledge to guide the choice of specific browsing and querying methods that are driven by specific users' requests in that domain. The issues encountered during video processing, such as features extraction, are addressed differently depending on the video topic and so video information analysis and video information extraction techniques can differ significantly from one domain to another (for example, extracting information from a video of a soccer game would differ significantly from extracting information from television news or conference recordings). For what concerns the domain specific frameworks, we mention COALA [Fatemi 2003] for TV news retrieval, ConKMeL [H. 2006] for e-learning knowledge management, and the soccer framework [Z. 2008] for soccer videos. We believe that Matterhorn [MAT], VIKEF [VIK] and the RKBexplorer [Glaser 2007, Glaser 2008] are three interesting frameworks approaches that have drawn our attention.

4.2.1 Matterhorn

Matterhorn is an end-to-end software application that supports the scheduling, the capture, the management, the encoding and the delivery of educational audio and video content. Matterhorn has four key components: Lecture Capture and Administration, Ingest and Processing, Distribution Management, and Engage Tools. Matterhorn is a free, open-source platform to support the management of educational audio and video content. Institutions can use Matterhorn to produce lecture recordings, to manage existing videos, to serve designated distribution channels, and to provide user interfaces aim at engaging students with educational videos. The first release, Matterhorn 1.0, provides the following features:

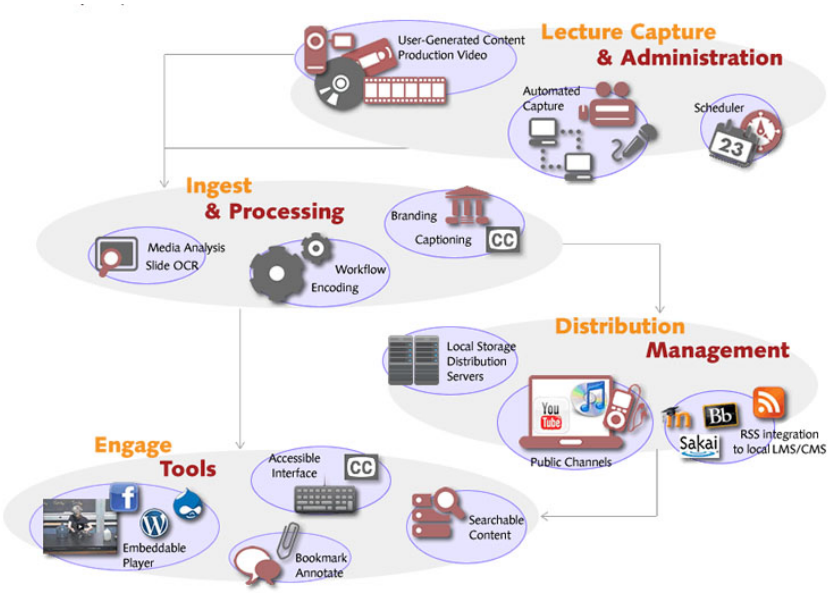


Figure 4.2: Matterhorn project

- Administrative tools for scheduling automated recordings, manually uploading files, and managing metadata, captioning and processing.
- Recommended capture agent hardware specifications.
- Integration with recording devices in the classroom for managing automated capture.
- Processing and encoding services that prepare and package media files according to configurable specifications.
- Distribution to local streaming and download servers and configuration capability for distribution to channels such as YouTube, iTunes and campus course and content management system.

- A rich media user interface for learners to engage with content, including slide preview, content-based search and captioning.

However, the Matterhorn framework has the following drawbacks: The Matterhorn project lacks semantic models that considerably enhance the information retrieval. Matterhorn does not offer users an ontology driven browsing and visualization approach. Additionally, Matterhorn does not address the issue of heterogeneous data.

4.2.2 VIKEF

The Virtual Information and Knowledge Environment Framework (VIKEF), is an application-oriented integrated project. VIKEF is dedicated to advanced semantic-enabled support for Information, Content, and Knowledge (ICK) production, acquisition, processing, annotation, sharing, and use by empowering information and knowledge environments for scientific and business communities. Within this context, VIKEF's main challenges focus on bridging the gap between the partially implicit knowledge and information that is conveyed in human produced scientific and business content resources and the explicit representation of knowledge required for targeted and effective access, dissemination, sharing, use, and annotation of ICK resources, both by the scientific and business communities and by their information and knowledge-based work processes. The main features of the project are:

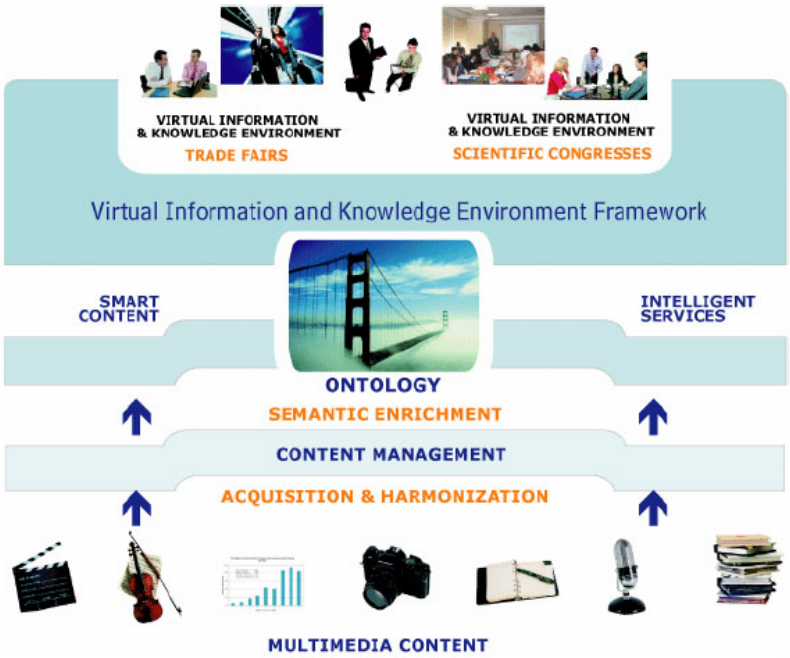


Figure 4.3: VIKEF architecture

- An advanced software framework for semantic-based Information, Content, and Knowledge (ICK) management systems and intelligent services supporting

Knowledge Value Chain (KVC) application solutions.

- An exemplary sectoral KVC application solution based on this framework providing semantic-enabled community services for the scientific and business domains.

The long term vision behind the project is that the VIKEF open and flexible framework will provide a wide spectrum of actors in different application domains with a robust architecture and software development environment for the rapid, flexible, and customized development of semantic-enabled virtual information and knowledge environments.

VIKEF presents an interesting approach that relies on semantic models. However this framework does not take into account the multimedia resources used in scientific events. It does not offer users ontology driven browsing and visualization approach. Finally, it does not address the issue of data heterogeneity.

4.2.3 RKBExplorer

RKBExplorer is a semantic web application that can present unified views of a significant number of heterogeneous data sources with respect to a given domain. The underlying information infrastructure is mediated by ontologies and consists of many independent triple stores (the subject, the properties and the object). RKBExplorer integrates a number of tools for both knowledge acquisition and publishing. It offers a faceted browser for navigating through information gathered

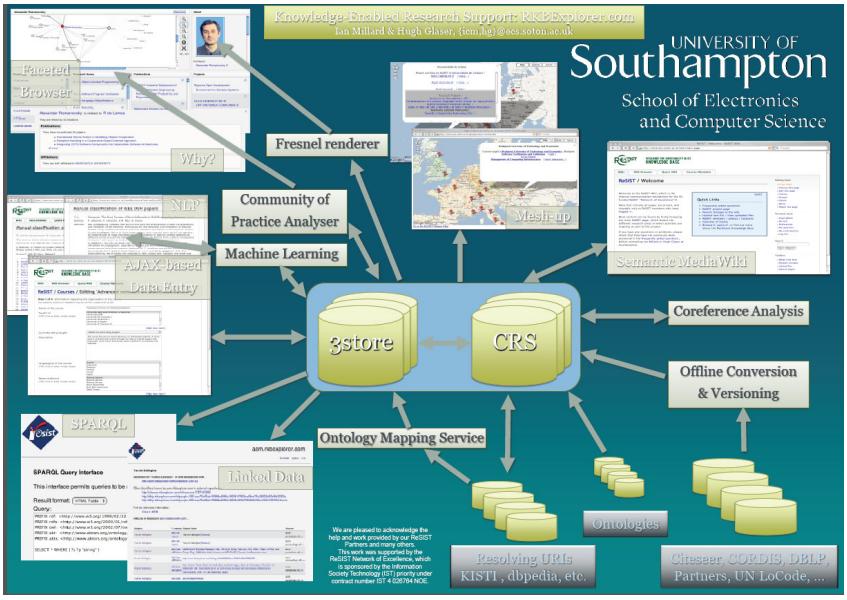


Figure 4.4: RKBExplorer approach

from a number of different types of resources, using techniques which seem most

appropriate for each type. Some techniques are simply web scraping, others such as DBLP, provide XML dumps which can be easily transformed into RDF. Users can navigate through publications, projects, and the scientific community.

RKBExplorer framework has the following drawbacks: It does not consider searching and navigating through multimedia information (including video recordings of scientific talks), and it does not take into account the information generated during conference life cycle.

4.3 P2P Networks

P2P systems are decentralized large-scale computer networks in which peers simultaneously operate as both clients and servers. Peers can join and leave the network at any time. The participants in a P2P network make a portion of their resources (such as processing power, disk storage or network bandwidth) directly available to other network participants. Peers are both suppliers and consumers of resources, in contrast to the traditional client server model where only servers supply and only clients consume. The P2P network architecture is commonly used for file sharing applications. [Lua 2005, Blanco 2006]. The P2P architecture is known as an overlay network architecture. Overlay networks are virtual networks build on top of one or several existing networks like the Internet: The latter networks are consequently, known as the underlay networks. The overlay network is formed by logically connected devices and entities belonging to the underlay network. P2P is a class of applications that takes advantage of distributed network resources to accomplish a specific function in a decentralized manner. As the name implies, in P2P networks the peers communicate and share their resources directly with one another without requiring central coordination. A P2P network is a self organizing system of equal autonomous entities, the peers.

The power of P2P networks lies in their ability to provide services with practically unlimited scalability, based on the principle of resource sharing. In the context of information retrieval, the principle of sharing also applies to the knowledge in document collections and to the retrieval models. By combining the resources of a very large number of peers, we can expect a qualitative shift in future information retrieval systems [Klampanos 2003]. Federated, decentralized search engines, in which the effort is shared among a very large number of computers, appear to offer the potential to build large information retrieval systems with almost unlimited capacity. There are several projects in the research community that use semantic P2P systems, namely the project Search In Audio Visual Content Using Peer-to-peer IR (SAPIR) [SAP], the Generic P2P Semantic Service Oriented Architecture (GSS) [Kilany 2008], Edutella [Nejdl 2002], GridVine [Cudre-Mauroux 2007], SEMantic Peer Layer InfrastruCturE (SEMPLICE) [SEM], RDFpeers [Cai 2004], SWOOKI [SWO] and DRAGO [DRA a] to cite few.

- The figure 4.6 presents the general application scenario of the project SAPIR: The sources of the multimedia data are crawled and indexed by the P2P-based system. The P2P system is composed of a set of normal and super peers who possess crawling, indexing, and searching capabilities.

- The figure 4.5 presents the architecture of a GSS peer which is based on the JXTA technology [JXT , Gong 2001]. The authors proposes a generic support of applications and the integration of a semantic layer at the resource description level.

- Edutella is a peer to peer network for searching semantic web metadata. Files are not actually shared in the network rather the information that describes what the data is about. Somewhat simplified the core of Edutella consists of a library and a query language. The library can be used to form new Edutella networks or build provider or consumer peers that connect to existing Edutella networks.

- GridVine is a Semantic Overlay Network supporting decentralized data integration techniques through pairwise schema mappings and monotonic schema inheritance.

- SEMPLICE is an extensible P2P Infrastructure for query and discovery across a semantic layer, based on domain and service ontologies, over the logical overlay network of a traditional P2P architecture.

- RDFpeers is a peer to peer RDF repository which stores each triple in a mutli-attribute addressable network by applying globally known hash functions. Queries can be efficiently routed to the nodes that store matching triples.

- SWOOKI is a peer to peer semantic wiki. Integration of semantic web technology allows building a structured knowledge base. While the peer to peer architectures provide decentralized, scalable and reliable services on top of a network of loosely coupled and weakly connected peers.

- DRAGO it addresses the problem of reasoning with multiple distributed ontologies interrelated by pair wise semantic mappings. From the implementation point of view GRAGO represents a peer to peer reasoning platform. It envisages a web of ontologies being distributed amongst peer to peer network of DRAGO Reasoning Peers.

As our intention is to enhance the multimedia information retrieval, we intend in this thesis work to make use of existing P2P solutions, in order to facilitate the resources and annotation sharing among the scientific community and to improve the accessibility of wider and distributed information. As stated above, in this thesis work we only make use of existing platforms and protocols and we have chosen the framework JXTA and the protocol Chord [Stoica 2003]. We chose these because

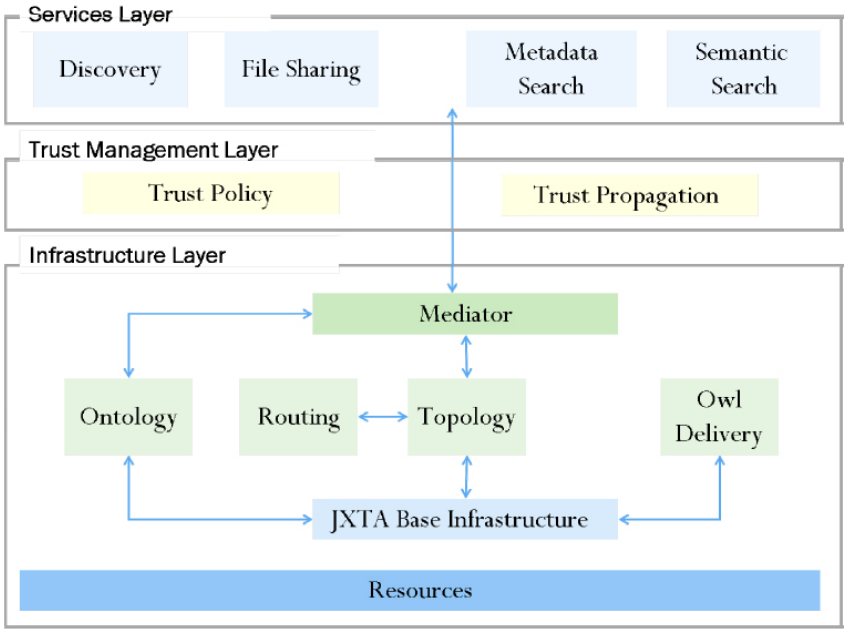


Figure 4.5: GSS architecture

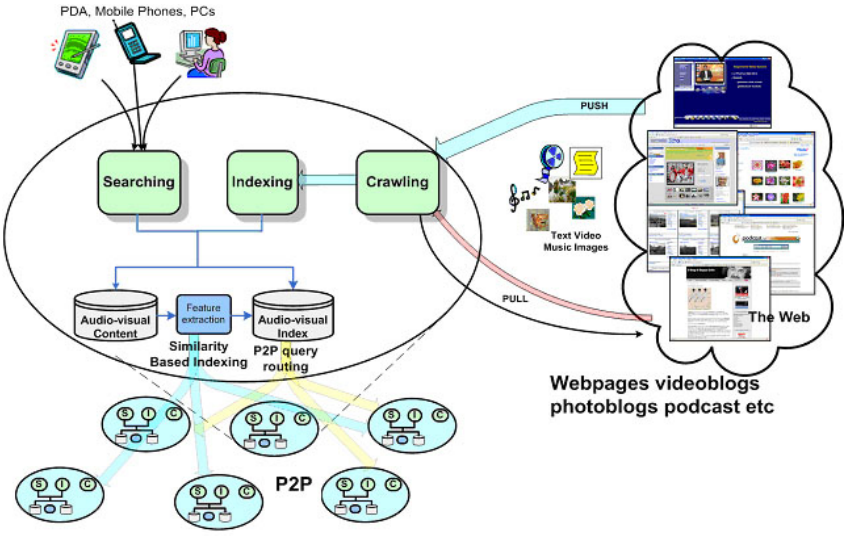


Figure 4.6: SAPIR approach

Chord and JXTA are both open source and they are easy to use and to implement. Adding to this, latest works that have been done over P2P networks were based on JXTA and/or Chord [P2P , Yeh 2010, Aberer 2005, Lorenzo Corsani] and have proven their effectiveness in this context. In addition JXTA supports the implementation of RDF-based metadata infrastructure for P2P networks, which is a key point for supporting semantic retrieval [EDU].

- JXTA Framework (Fig. 4.7): The Java JXTA platform is a series of classes and methods for managing and transmitting application and control data between JXTA compatible peer platforms. One of the advantages of JXTA is the portability of the P2P network implementation. This platform uses six protocols to create P2P applications: Peer Discovery Protocol, Peer Resolver Protocol, Peer Information Protocol, Peer Membership Protocol, Pipe Binding Protocol and Peer Endpoint Protocol. Each peer in a JXTA network belongs to one or more groups and implements a set of services that allows interaction with other peers. For most applications, the peer is synonymous with the user. A JXTA peer in the Java implementation is associated with one Java Virtual Machine (JVM). Their is usually one peer per device, such as a PC. For debugging and experimentation, we can start multiple JVMs on the same peer. In this project, we use the VMware tool that allows us to run a virtual machine on a real one, thus allowing us to simulate the P2P network with only one JVM on each machine using separate directories. The JXTA protocols are based on XML messages. These XML messages are passed between the peers to convey information or are exchanged as part of a longer communication with queries and responses. The protocol defines the format of the XML messages and the rules under which they are sent.

- Chord [Stoica 2003]: Chord is a distributed lookup protocol that addresses the problem of efficiently locating the node that stores a particular data item. Chord supports mainly one operation: Given a key, it maps the key onto a node. Depending on the application using Chord, that node could be responsible for storing a value associated with the key. Each node maintains information of a limited number of other nodes and resolves all lookups via the same number of messages to other nodes. Chord maintains its routing information as nodes join and leave the system. Three features that distinguish Chord from many other P2P lookup protocols are its simplicity, provable correctness, and provable performance. Using Chord it is simple to route a key through a sequence of other nodes toward the destination. A Chord node requires information about other nodes for efficient routing, but performance degrades gracefully when that information is out of date. Only one piece of information per node needs to be correct in order for Chord to guarantee correct routing queries and it has a simple algorithm for maintaining this information in a dynamic environment. It does not provide anonymity, but its lookup operation runs in predictable time and always results in success or definitive

failure. It has the advantage that its correctness is robust in face of partially incorrect routing information. Chord simplifies the design of P2P systems and applications based on it by addressing these difficult problems:

- Load balancing: Chord acts as a distributed hash function, spreading keys evenly over the nodes; this provides a degree of natural load balance.
- Decentralization: Chord is fully distributed; no node is more important than the any other. This improves robustness and makes Chord appropriate for loosely-organized P2P applications.
- Scalability: The cost of a Chord lookup grows as the log of the number of nodes, so even very large systems are feasible. No parameter tuning is required to achieve this scalability.
- Availability: Chord automatically adjusts its internal tables to reflect newly joined nodes as well as nodes failure, ensuring that, barring major failures in the underlying network, the node responsible for a key can always be found. This is true even if the system is in a continuous state of change.
- Flexible naming: Chord places no constraints on the structure of the keys it looks up. The Chord Key-space is flat. This gives applications a large amount of flexibility in how they map their own names.

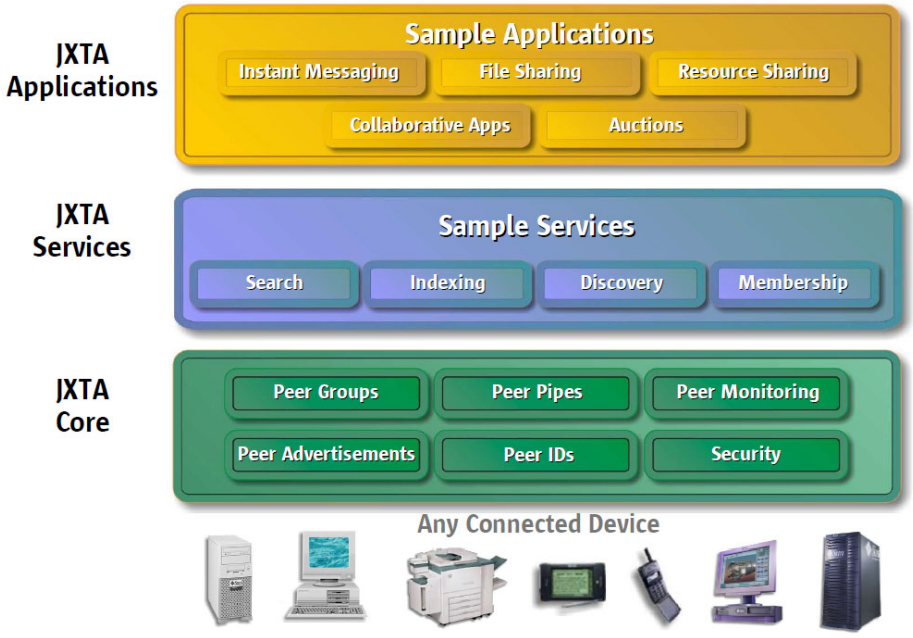


Figure 4.7: JXTA software architecture

4.4 Conclusion

The ever increasing growth of available information makes finding pertinent information a challenging one, calling for novel approaches. Traditional data and information retrieval systems supporting current web information are no longer adequate for knowledge retrieval tasks. The knowledge retrieval systems will be the next generation's information retrieval systems. A prominent conclusion of this study is that domain based frameworks are very important in this respect. Domain specific approaches rely on specific domain knowledge, thereby enhancing high-level information extraction and enabling the design of accurate tools and algorithms.

The study shows that few frameworks make use of the potential offered by P2P architectures. Many P2P file sharing applications exist but they lack major feature: Semantic search. While most P2P sharing applications offer keyword based search in data indexes, there are few that really access the content of data and make use of powerful annotation schemes based on semantic techniques (RDF, OWL, and the like), or that offer content based information about the data.

The thesis proposed approach promotes the use of an integrated end-to-end framework that integrates novel approaches to multimedia information retrieval, namely, the annotation, browsing, and visualization approaches and that is based on a P2P network in order to facilitate sharing among the scientific community and to improve the accessibility of wider and distributed information.

Chapter 5

Discussion

The overall analysis of the state of the art presented in this part led us to conclude with a set of observations which motivate the main research axes of the work proposed in this thesis. The main issues addressed in this thesis work are relative to the CBMR process issues (Fig. 3.1), they include the four points:

Knowledge modeling: The design of a semantic model for the conference multimedia information retrieval process. In order to support a general purpose multimedia retrieval system, such as scientific recordings, the model should combine rich set of various structural and content based descriptions. In this context the use of ontology may be of great interest. The study shows that an increasing attention is given to ontology based approach, as it offers promising solutions to enhance and allow information retrieval at a semantic level. Indeed, when ontology techniques are coupled with information retrieval tools, they provide an inferring process which make possible accurate interpretation of the information and therefore efficient and pertinent retrieval of such information. Existing models cover some of the required structures and content based descriptions. In the chapter 6 of this thesis we provide the knowledge model, HELO, which is based on and extends existing ontologies.

Knowledge management: The improvement of the content based multimedia description by aggregating available information. The ones generated through the chain process of the data production, the automatically extracted information and the manually annotated multimedia data. Few frameworks implemented these different approaches all together because of the heterogeneity of the resulting formats. This heterogeneity calls for an effort at the querying level. The chapter 7 (section 7.1) presents the knowledge management approach that makes use of the information that we may collect through the life cycle of a multimedia data, regardless of the description format .

Knowledge browsing and visualization: The improvement of the content based multimedia retrieval (by offering model based browsing approach) and the results visualization. Using the semantic technologies may have a significant potential to improve tasks related to multimedia retrieval by making the multimedia resources more accessible, understandable, sharable and dynamic. Still, to be effective query interfaces and the results rendering have to be easy to read and simple to navigate through. The section 8.1 presents the querying approach adopted in this thesis work that addresses the issue of data heterogeneity. In the section 8.2,

we present the knowledge browsing and visualization approach based on domain specific model that take into account the life cycle of information retrieval.

Conference multimedia framework: The design of an integrated end-to-end framework for the conference multimedia retrieval. The framework unifies the many efforts that have been conducted so far and go further by integrating novel approaches within the content based multimedia retrieval process, namely the annotation, the browsing and the visualization. Currently existing content based multimedia retrieval frameworks do not make use of the life cycle of a multimedia resources, few of them lie on a semantic model for the entire information retrieval process. Moreover, none of the domain specific frameworks presented in the state of the art benefit from the potential that P2P networks may offer. The chapter 9, presents the ontology driven framework based on the P2P network.

The table 5.1 presents a comparison between the main works presented in the state of the art. We cite in the first row of this table the works that represent the greater number of similarities relative to this thesis work. We set up the comparison according to the addressed issues over a Content Based Multimedia Retrieval process namely the Multimedia Modeling, the Multimedia Management and the Multimedia Searching and Visualization (Fig. 3.1). For each of these issues we put forward relative aspects which we considered specific and significant for the design of a content based multimedia retrieval framework in the domain of scientific conferences: (1) Models Designing: We aimed at identifying the works that concern the conference domain, the ones that modeled the knowledge conveyed within this realm and the works that addressed the multimedia aspects more precisely the modeling the videos. (2) Multimedia Management: We aimed at identifying the works that took advantage of the information conveyed within a conference life cycle, the works that addressed the ontology driven knowledge management, the works that focused on the synchronization of the different type of multimedia resources (video, presentation file, annotation, and so one) and finally we focused on the works that used an hybrid approach (automatic, manual and semi-manual approaches) for the knowledge management. (3) Multimedia Browsing and Visualization: We aimed at identifying the works that were done for the multimedia browsing and visualization by focusing on those that were driven by an knowledge model. (4) Integrated Framework: As for the design of the framework, we focused on the analysis of the tools independent frameworks, the ones that were based on a P2P network, the ones that allow the integration of different metadata formats.

Based on this overview (Tab. 5.1), we identify the different thesis contribution issues that are addressed in order to offer a novel approach of an ontology driven framework for a conference multimedia retrieval in a P2P network.

Issues	Sub-issues	ISWC [Knud Moller 2007]	VIDEOLECTURES [VID b]	VIKEF [VIK]	MATTERHORN [MAT]	SAPIR [SAP]	RKBEXPLORER [Glaser 2007]	SEMANTICLIFE [Hoang 2006]	DBLPVIZ [DBL a]	ARNETMINER [ARN]	**THESIS CONTRIBUTION** CALIMERA [Sokhn 2010b, Sokhn 2011b]
Models Designing	Conference Domain	■	■	■	■		■		■	■	■
	Conference Knowledge Modeling	■					■				■
	Video Conference Ontology Model										■
Multimedia Management	Conference Life Cycle Knowledge Management				■						■
	Ontology Driven Knowledge Management			■			■				■
	Multimedia Synchronization				■						■
	Hybrid Knowledge Management				■						■
Multimedia Browsing and Visualization	Multimedia Browsing and Visualization		■	■		■	■			■	■
	Knowledge Driven Browsing and Visualization			■							■
Integrated Framework	Tools Independent Framework				■						■
	Usage of P2P Potential for Multimedia Retrieval					■					■
	Multi-format Metadata							■			■

Figure 5.1: Overview of the main works presented in the state of the art

Part III

CONTRIBUTION - Conference Multimedia Information Retrieval

Chapter 6

Knowledge Modeling

6.1 Introduction

One of the most important conclusion that comes up from the study of the state of the art, is the need for an end-to-end integrated framework guided by a knowledge model designed for the conferences domain. Based on this observation and by analyzing the conference life cycle we provide an integrated approach to the retrieval and visualization of the conference multimedia resources based on a conference knowledge model. As stated earlier, the term 'knowledge' according to Oxford Dictionaries Online means "Facts, information, and skills acquired through experience or education; the theoretical or practical understanding of a subject'. In the rest of this report the term 'knowledge' is used to designate an understanding of a specific domain that goes beyond the simple information and integrates aspects such as competences, skills, expertise, etc. In this chapter we present the conference life cycle notion and we describe the conference model.

6.2 Conference Life Cycle

Every conference event may be divided into several differing steps. Each of which conveys a set of information that can be used in the information retrieval process. A conference life cycle is composed of this series of differing steps (Fig. 6.1):

- **Conference Pre-Event**: This step is relative to the conference organization issue, mainly the administrative part. In this step, the organizing committee prepares the conference program, the venue and accommodation, etc., the participants register at the conference adding their name, affiliation, etc., the keynote speakers provide a short abstract of their talk and a small biography, etc. This information can be acquired and linked to the video recordings, generated at the next step, the Conference Event, in order to enrich their description.

- **Conference Event**: During the conference event, most of the talks are recorded as well as the slide-show presentations that may also be acquired during the event. In this step, the main data resources are captured. Additional resources may be also captured after the talk event within the Post-Event Data Oriented step.

- **Conference Post-Event**:
  Data Oriented - this step is relative to the acquisition of different data resources

related to the talks such as the related articles, video of demonstrators, photos of the event, information about related event, etc. Some peers may share their data and enlarge hence the resources collection
Metadata Oriented - in this step, some resources may be annotated to enrich the original description of the media such as the manual annotation of the video sequences, some other information may be automatically extracted such as the temporal information relative to the slide-change detection of the talk video recording. This information, in addition to the one presented in the previous steps are intended to be published for later retrieval.
Publication Oriented - at this level of the conference life cycle, the resources (local to a system or shared through a peer community) and their descriptions are available to be published. This multimedia information (video, image, text, etc.) is made accessible through a search engine.
Resources Accessibility - at this level, the users intend to access the resources throughout a search engine, the users may look up for these resources (locally or on a peer) through a browsing and/or a querying process.

The retrieval of a talk video recording may be considerably enhanced if the information generated through a conference life cycle is captured. In order to support an effective management and reuse of the knowledge conveyed within the conference life cycle we proposed an ontological model. This model, described in the following section integrates the information related to a conference life cycle, hence, enriching the description related to a video talk recording.

6.3 HELO Model

In their reviewing [Snoek 2005] Cees S. et al. present three issues arising when assigning an index to a video. The first one is related to the granularity, it addresses the question: **What to index**, e.g. the entire document or a single frame. The second is related to the modalities (audio, image, etc.) and their analysis, it addresses the question: **How to index**, e.g. a statistical pattern classifier applied to the auditory content only. The third is related to the type of index used for labeling, it addresses the question: **Which index**, e.g. the names of the presenters in a conference. In this thesis work we address a forth issue: **Why to index** or the **user dependent indexing**. It addresses the requirements of users and their behavior while searching for multimedia information. In this context, HELO has been designed, aiming to enhance the multimedia information retrieval, more precisely the scientific conference recordings retrieval.

HELO is an ontological model. In fact, the use of ontologies makes possible the extension of existing vocabularies and concepts. Moreover the data and object properties such as symmetry, transitivity, inverse functional, etc. simplify and enhance the knowledge modeling and its uses. As discussed in the state of the art section, several effort has been made to model the conference context. We based

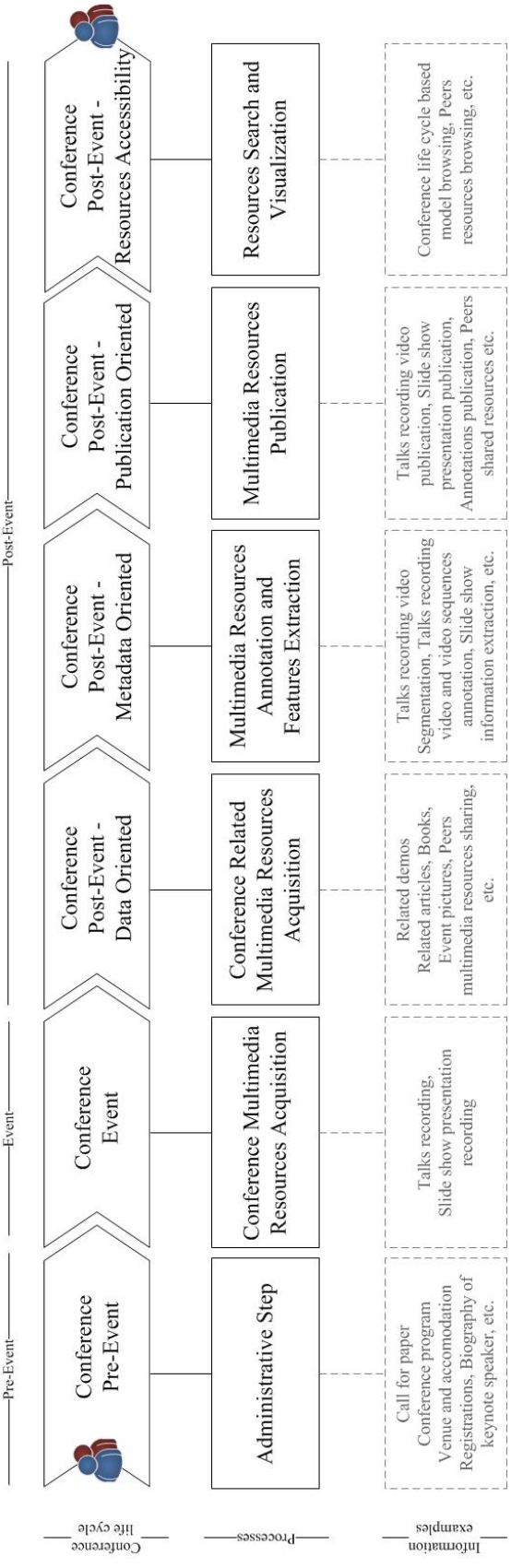


Figure 6.1: Conference life cycle steps description

this thesis work mainly on the effort done by "The ESWC and ISWC Metadata Projects" since they proved their usability as stated in the chapter 3.

HELO describes and structures the information conveyed within a conference life cycle. The knowledge conveyed along a scientific conference life cycle concerns mainly three entities:

- **Event** - related to Conference Pre-Event and Event step: Events may consist on a single talk such as a paper presentation, or multiple talks within a session, or an industrial track or special events such as an inauguration or a software release, and so on.
- **Artefact** - related to Conference Event and Post-Event Data Oriented steps: Artefacts are about multimedia data such as the talks recordings, full papers or posters, conference proceedings, speaker's slide set, and most importantly and of great value -according to this thesis work, the videos of the recorded talks.
- **People** - related to Conference Pre-Event and Post-Event Metadata Oriented steps : People are about speakers, community, attendees, organizers, and more.

Wherever possible we made use of established concepts. HELO uses mainly the ESWC2006, ISWC2006 ontologies. It integrates several other concepts extracted from other established ontologies related to conference and other broad domains. It makes use of the Event concept from the SWC ontology [SWC], the Person concept from the FOAF ontology [FOA], the ResearchTopic concept from the SWRC ontology [SWR]. In addition to the reused concepts HELO developed additional concepts/classes mainly for modeling the multimedia resources information: Talk video recordings, presentation slide-show and other data resources related to the talk. HELO can be graphically presented as a two-set of concepts (Fig. 6.2): The **annotation oriented set** and the **user oriented browsing & visualization set**.

Based on this, the knowledge conveyed within a conference may be on the one hand described and annotated according to granular concepts (**annotation oriented set**) and on the other hand it may be searched based on a user oriented set of concepts (**user oriented browsing & visualization set**):

The **annotation oriented set**: It offers granular concepts allowing users to annotate conference information, and mainly scientific talk recordings, with a high-level degree of expressiveness. This set is composed of the following concepts: Group, Person, OrganisedEvent, Location, MultimediaDocument and Topic. Other important concepts have been defined as sub-concepts/sub-classes or properties such as the concepts of *VideoRecordings*, *Role* and *Expertise*. The major improvements of HELO compared to the listed ontologies are the following:

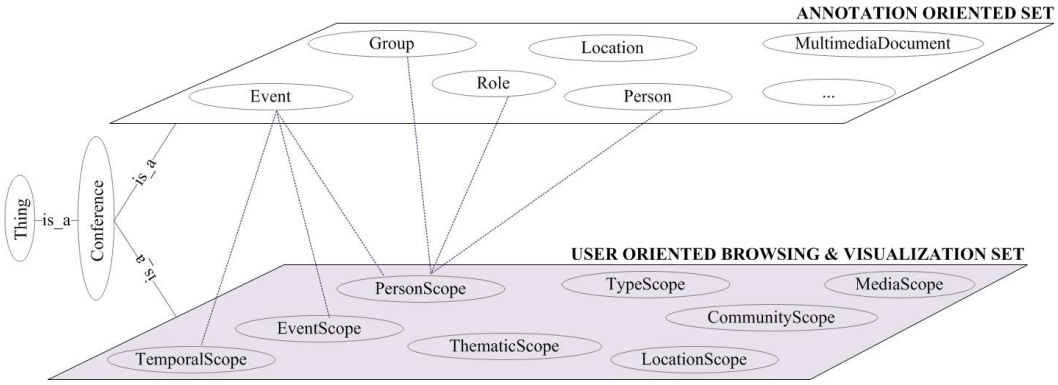


Figure 6.2: HELO - conference model overview

- The ***MultimediaDocument*** concept which describes all the media resources conveyed within a scientific conference.
- The ***VideoRecordings*** (a sub-class of *MultimediaDocument*) which describes the video talk recordings of a scientific conference.
- The ***Expertise*** concept which describes the expertise level per domain.
- The ***Topic*** concept which describes the thematics occurring in scientific conferences. It is based on the ACM taxonomy [ACM].

The **user oriented browsing and visualization set**: It offers the users a faceted search and visualization interface. The different facets are being referenced in this thesis work as *scopes*. Searching for a video talk recording of a conference may be performed in different ways (by speaker, by event, by topic, etc.). We propose eight *scopes* each of which represents a concept in the proposed ontology and is related to one or more concepts of the *annotation oriented set*:

- ***PersonScope*** includes information about people involved in a conference e.g. names, roles, affiliation. This information may come from the Conference Pre-Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find the video-recording of the talk where a colleague of the chairman made a presentation. It allows users to search by presenter, audience, chairman, committee and so on.
- ***LocationScope*** contains information about the conference location e.g. continent, city, building, room. This information may come from the Conference Pre-Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find the video-recording of the talk that took place in a city in France. It allows to search for example where the conferences took place.
- ***TemporalScope*** concerns conference planning e.g. starting time, parallel sessions, breaks. It allows users to make queries such as: Find the video-

recording of the talk that took place in the afternoon in parallel with the talk B.

- ***TypeScope*** lists several categories of conferences e.g. workshop, lecture. This information may come from the Conference Pre-Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find a talk given in the academic lecture training.

- ***MediaScope*** gathers all the media information linked to a talk e.g. video-recording of the talk, slide set document, papers, books. This information may come from the Conference Event and Post-Event Data Oriented steps. It allows users to make queries such as: Find the video-recording sequence of the talk related to this paper. This scope interests us particularly due to the rich information we can extract from each implied media, hence increasing the semantics level. e.g. we can extract information from the Presenter Talk, to distinguish introduction part from conclusion, demo, questions, and the like.

- ***ThematicScope*** affiliates a conference to a domain, topic, related events e.g. video-recording indexing, biology. This information may come from the Conference Pre-Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find the video-recording part of the talk related to knowledge management. It also assigns to a conference an abstract, some keywords and related events. Thus we give to the users the possibility to search by topics rather than only by keywords. Users could be interested in finding a conference related to an event as an inauguration or a product release.

- ***CommunityScope*** defines communities such as laboratories, research groups, conferences committees e.g. MEDES program committee, MIT group. This information may come from the Conference Pre-Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find the video-recording of the talk where a professor from France university in the MEDES program committee made a presentation. Using this information a user can search a conference given by a professors' team or any person in his/her laboratory.

- ***EventScope*** describes the events related to a conference event. This information may come from the Conference Pre-Event, Event and Post-Event Metadata Oriented steps. It allows users to make queries such as: Find the video-recording of the talk related to the Large Hadron Collider (LHC) inauguration event.

The figure 6.3 shows the kinds of MultimediaDocument elements that have been added in this thesis work and that can be said about each of the main entities (represented in purple) surrounding a paper which are: The video recording related to the paper presentation - *Helo:VideoRecordings*, the video segments that are related to the video recordings - *Helo:VideoSegments*, and the slides extracted

from the file of the paper presentation - *Helo:VideoSegmentsSlide*. Each of the listed entities is related to other entities (represented in white) or values. For instance the *VideoSegments:79_5 hasDuration 00:01:58* and is sameAs the entity *video_segment_demo:795_5*. Wider extract of the model HELO are given in the appendix A.

By using the proposed model HELO to index, we offer users the possibility to perform complex requests such as "*Find me an audiovisual sequence* (***MediaScope*** involved) *in an academic training lecture* (***TypeScope*** involved), *in 2007* (***TemporalScope*** involved), *in Italy* (***LocationScope*** implied), *where a colleague of a professor* (***PersonScope***, Committee involved) *talked about muon* (***ThematicScope*** involved) *just after the coffee break* (***TemporalScope*** involved)".

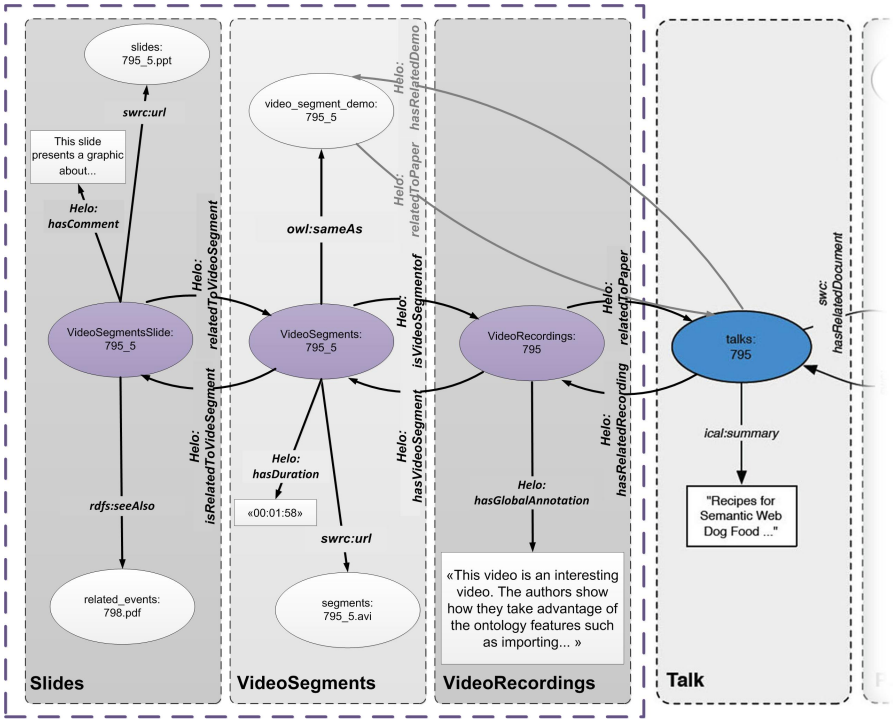


Figure 6.3: HELO - part of the thesis proposed model

6.4 Conclusion

We presented in this chapter the model HELO based on the information that we can find along a conference life cycle. This information provided at the Pre-Event, the Event and the Post-Event steps may considerably enhance the multimedia information retrieval, more precisely the conference talk recordings retrieval. The enhancement arises along the retrieval process namely the browsing of the published

resources and the visualization of the rendered result. Capturing this information may be a tedious task. We aim in this thesis work to enhance this task by combining the different approaches (automatic, manual and semi-manual) for knowledge management. In the following chapter we present the model based approach used on the one hand for the knowledge management and on the other for the knowledge based searching mainly the browsing and visualization aspects.

Chapter 7

Knowledge Management

7.1 Introduction

As described in the previous chapter, a conference life cycle conveys a set of information that may, if captured, enhance the multimedia information retrieval process. In order to use this information we designed HELO, a conference knowledge model. This model is used during the management and the searching phases. In this chapter we present the thesis proposed approach for the content based multimedia management. We present the metadata extracted automatically (Automatic Content Extraction) and the ones added manually or semi-manually (Manual and Semi-manually Content Based Annotation).

7.2 Multimedia Knowledge Management

Different types of information can be associated to a multimedia resource such as a video. Some of them can be automatically extracted, others have to be manually integrated into the video description by an annotation process. In this thesis work we aim at enabling the creation of semantic based knowledge management approach helping to bridge the so-called "semantic gap". The main target of this thesis work was focused toward making use of information generated through a conference life cycle that may be scattered because of the diversity of isolated digital realms. To this respect we worked on the integration, extension and improvement of existing works based on automatic and semi-manual information management process. We identify three different approaches: The automatic content extraction, the manual content based annotation and the semi-manual content aggregation.

7.2.1 Automatic Content Extraction

Extracted content can be categorized based on the low-level features such as the color, the shape, and so on, and the high-level ones such as keywords, face recognition and the like. Automatic content extraction is based on techniques such as image or speech analysis techniques. In this section we present two of the approaches that we decided to investigate in this thesis work: The content based video segmentation approach based on a low-level feature extraction techniques and the high-level content extraction based on a speech extraction techniques.

7.2.1.1 Content Based Video Segmentation

This section presents the thesis proposed algorithm that automatically segments the video recording of the presentation and aligns each segment to the corresponding slide. In most case, in scientific talks, the speakers make use of slide-based presentations (ie. Microsoft PowerPoint files) displayed as a slide-show during the speech. Data contained within these electronic documents can be used to improve video recording indexing in order to facilitate retrieval of specific sequences within the whole video recording. In order to exploit such data it is however necessary to synchronize the conference talks recording (video) with the corresponding slide-show presentation. So far such synchronization has been done mainly manually. Nowadays, given the large amount of conferences being recorded, manual archiving is becoming a too time-consuming task that need to be automated.

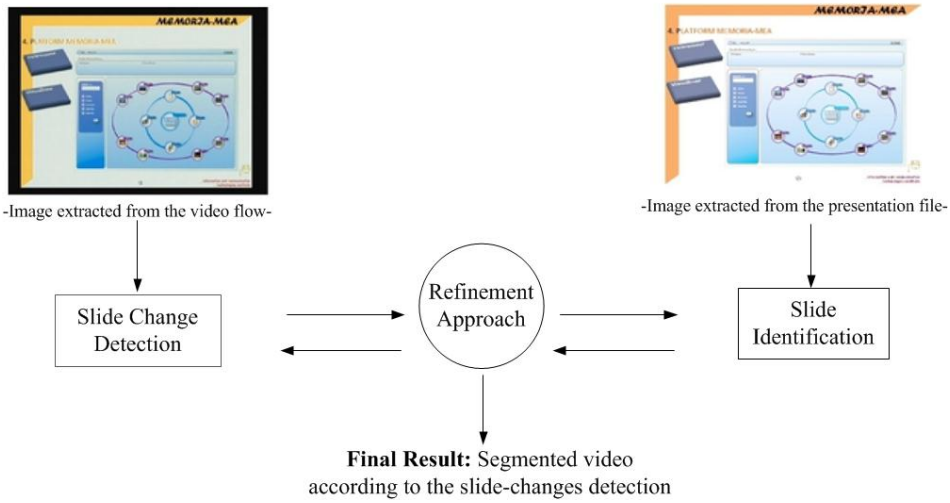


Figure 7.1: Overview of the slide-changes based video segmentation approach

Algorithm concept:
The figure 7.1 presents an global overview of the content based video segmentation approach proposed for the slide-video synchronization. The algorithm approach is described below. It is composed of two main phases process 7.2. The first phase consists on the segmentation of the video of the slide-show according to the slide-changes during the talk presentation. In the second phase, a key frame is assigned to every video segment detected in the first phase, and than this key frame is assigned to an image extracted from the presentation file. A refinement approach is performed over these phases taking into account heuristic aspects such as the fact that in most cases, the slides are shown sequentially from the first to the last slide.

Phase 1 - segmentation:

The first phase of the algorithm detects the slide transitions (slide-change) among the video frames in order to segment the video. Since the video of a slide-show displays mostly static patterns (ie. with a lot of redundant frames) the synchronization is performed on a per-segment basis rather than trying to match each frame individually. The perfect result being that there is exactly one segment for every occurrence of a given slide. Such transition detection system can be considered as a specific video shot boundary detector. Video shot boundary detection is a mature field in video analysis and includes several available algorithms, as the ones stated in [Boreczky 1998, Lienhart 1996, Yuan 2007]. However, the cited algorithms are designed for generic video sequences and not for slide-shows, which are mainly composed of static frames, with quick hard cuts, involving only a small number of pixel changes (all slides in a presentation generally present a very homogeneous style). Thus it is necessary to develope a specialized system [Wang 2003].

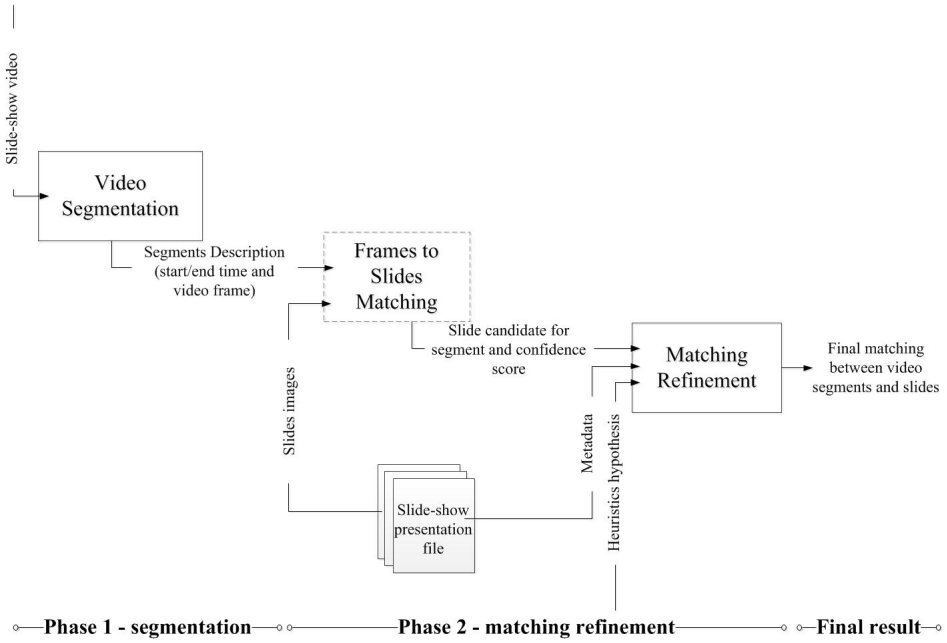


Figure 7.2: Slide-change based video segmentation process

In such systems, false-positive (spurious detection) will often occur due to animations in the displayed slide causing activity on the screen. On the other hand, if the sensibility of the system is not high enough, false-negative (slide-changes not detected), may also occurs. False-positives result in redundant or non-identifiable video frame, which could usually be filtered in a second time (ie. when proceeding with the identification of frames). Conversely, false-negatives are nearly impossible to fix, since they are very difficult to detect. Therefore the algorithm has to be sensitive enough to remove a maximum of false-negative, even if it does imply more false-positives [Mukhopadhyay 1999]. We proceed as follows: First, the absolute difference between two adjacent frames is computed and the resulting

values are stored into three 2D-matrix, one for each color channel. Then we calculate the variance of each matrix (Fig. 7.3). The variance provide a more convenient value for the thresholding, and allow us to ignore uniform alteration between two frames which may be due to noises, these three values are then summed and compared to a pre-defined threshold: If the result is greater than the threshold, a transition is detected. This threshold characterizes the sensitivity of the proposed system. In order to avoid an overload of transitions detection (e.g. when a video is running on the screen or during an off-slide demonstration), we stop storing video frames when an excessive number of transitions has been detected in a slot of time and we start capturing frame again as soon as a stable state is retrieved. Finally, for each found segment, a significant key-frame has to be chosen in order to be compared later with the slide pictures. It has been decided to simply take the last one, since it is the more likely representing a slide in its final stage.

Boundary shot detected only if: $Var(M_r) + Var(M_g) + Var(M_b) > change\_threshold$

Figure 7.3: Slide-change based video segmentation - boundary shot detection

Phase 2a - matching refinement:
Frames identification: Reviewed papers mostly presented their own solution to the issue of image comparison. Some of them used Discrete Cosine Transform (DCT) [Chiu 2000], other Scale Invariant Feature Transform (SIFT) [Fan 2006], while some other tried to use combinations of various features [Behera 2004]. All of them presented fairly good results. We choose an algorithm which has already been proved successful in a case of use close to ours: An edge histogram detection specialized for text content [Erol 2003].

Identification refinement: In most cases, the slide-show video ($>$ 60% in some cases) could represent frames which do not correspond to the pictures we have extracted from the presentation file. This is due to animations or external demonstration. This gives places to a large amount of segments where only few of them will have an usable identification. Once the video segmentation is available as well as the features vectors from the significant video frames, and the ones from the slide pictures extracted from the presentation file, the following steps start:

- Comparisons of features vectors: The system compares each frame features

vector with all the slides features vectors. The results are stored into an internal data structure.

- Deduction of candidates: All segments have at least one slide candidate, the one producing the best (lowest) difference score when compared with a segment frame (Fig. 7.4). When the result of the identification is ambiguous, additional candidates could be kept by inducing a concept of *relative confidence* of the identification. In order to be kept, the i-th additional candidate (starting from the second best matching slide to the worst one) have to have a relative confidence C_i greater than a given threshold. With C_i being:

$$C_i = \frac{s_i - s_0}{s_0} \tag{7.1}$$

  With S_i being the difference score of the i-th candidate, and S_0 the difference score of the best candidate. Segments providing a high confidence identification will likely have a single candidate for identification (the best one). The figure 7.5 presents an example of potentially matching slides candidates.

- Filtering non-usable frames: Most probably, some selected frames do not match with any slides (e.g. during external demonstration), they will likely present a low-confidence with many available candidates (ie. the difference scores are uniformly distributed, without strongly advantaging any slide).

- Deduction of the base sequence: The key heuristic employed here, is to consider that slides are sequentially displayed from the first slide to the last one. The chosen segments are ordered sequentially in time, the identified slides numbers have to be sequential with respect to the segment order, and the sum of the concerned identifications confidence has to be maximized. This concept of base-sequence provide two interesting features: (1) If an identification was unsuccessful, we still have a chance to get the right candidate kept (ie. not forcibly the candidate which had the lowest difference score), (2) the members of the base sequence are identified with a higher reliability: So a future occurrence of a frame identification result should have a features vector quite similar.

- Add non-sequential displayed slides: Non-sequential slides occur when the presenter goes backward in the slide stream, (and are therefore not contained in the base sequence). A non-sequential slide is added to the intermediary result only if the matching has a good confidence and if the sequence last more than five seconds. Adjacent sequences assigned to a same slide are merged.

- Add remnant slides: Remnant slides are non-sequential slides which do not fulfill the condition listed in the previous step. In practice, they only need a lower confidence and don't have to last a minimal amount of time to be kept, but they matched the same slide id as an adjacent already assigned sequence. This could happen when a slide is in an intermediary state and provides a

non-perfect identification (ie. with a low confidence). Adjacent segment with redundant slide assigned are merged.

- Assign orphan sequences: Addressed in the next section.

Phase 2b - matching refinement with orphan sequences assignment:
There is some case where a video segment could not be matched to a slide, this may be due to the animations displayed, or because of an external demonstration (e.g. opening a web browser) occurred. These non-identifiable sequences which remain after the first phase of matching refinement are called orphan sequences. As described above, a segment matching is likely to be succeeded when a slide is displayed in its final state. This allows us to define a default behavior when in presence of orphan sequences: Most of time they have to be assigned to the same slide than their followings, formally identified, neighbor. This default behavior already works well in the majority of the cases. It can be further enhanced by extracting internal information related to the slide-show presentation such as the identification of animated slides. Based on this we propose a set of rules:

- The previous and the following identified sequences are static slides. The orphan sequence is very likely an off-slide demonstration, and most probably linked to the previous sequence (a demonstration is almost always related to the current slide, which may sometimes be just an introduction to the demo itself): Assignment to Previous Sequence (PS).

- The previous sequence is static and the following is not. The orphan sequence could be present due to an off-slide demonstration, or an animation related to the following sequence. External demonstrations being naturally less frequent than animations: Assignment to Following Sequence (FS).

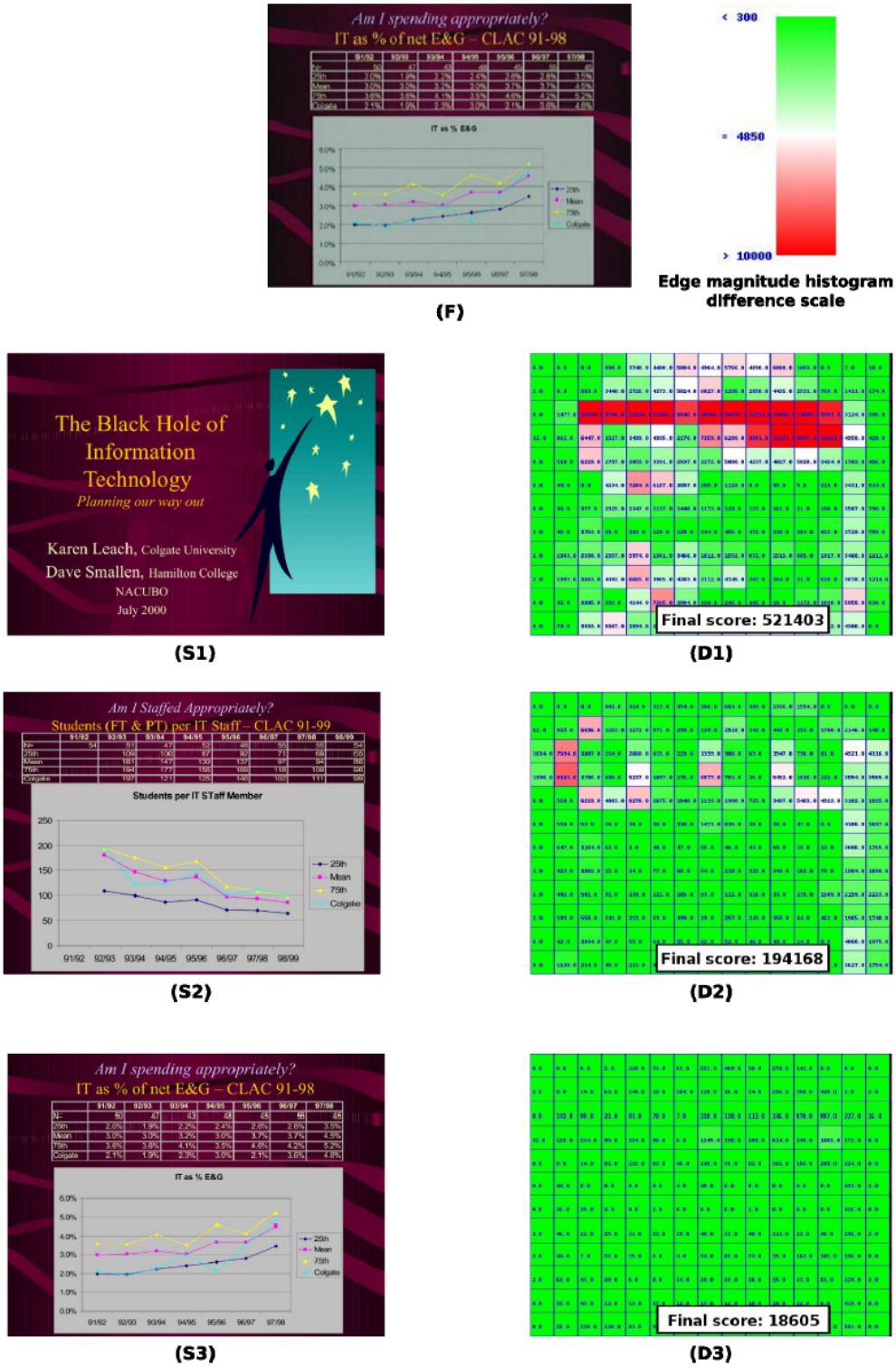
- The following sequence is static and the previous is not. The orphan sequence presence is due to an animation in the previous sequence or an external demonstration. In both case of figure: Assignment to PS.

- The previous and the following identified sequences are non-static slides. The orphan sequence presence can be due to animations either in previous or following sequences, or to an off-slide demonstration. We fall back to default most likely behavior (identifications succeeded with slides in their final states): Assignment to FS.

The segmentation of the video recording talks based on their content is an automatic process. This segmentation process can be coupled to the content based description process (manual or automatic annotation) allowing granular annotation via the possibility of describing video sequences.

7.2.1.2 Speech Based Metadata Extraction

The speech-to-text transcription can further improve the video recording indexing by assigning keywords, provided by the speech-to-text analysis algorithm, to a



(F): the video frame compared against the slides. (S1): a slide with no particular similarities with (F). (S2): a slide quite similar to (F). (S3): the slide perfectly matching (F).

Figure 7.4: Comparison of a video frame against the images of the presentation slides

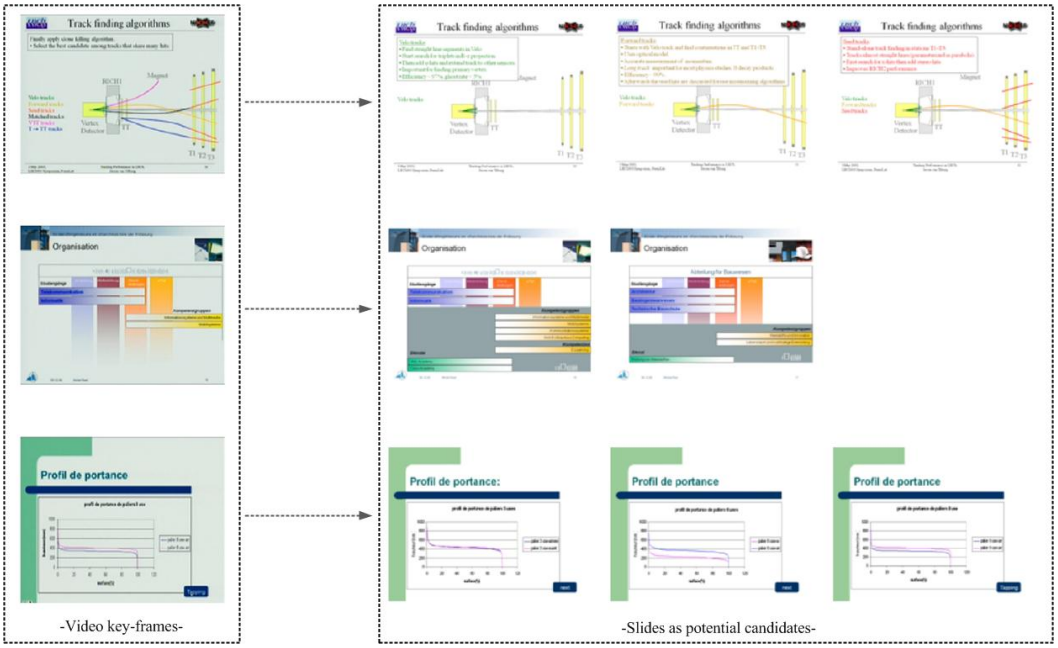


Figure 7.5: Video key-frames and their potentially slides candidates

video or video sequences. In order to exploit this information we developed an algorithm presented in this section. It automatically extracts time-based metadata keywords and link them to the corresponding video sequence.

In order to analyze the audio content of the video, a first step is to transcript the spoken words. The voice recognition technology uses methods from the signal processing and artificial intelligence domains. A recorded and digitized sentence is pushed to the voice recognition software in order to be transcripted. There are several voice recognition software solutions, among the best we can list Crescendo [CRE], Dragon Naturally Speaking [DRA b] and Sphinx4 [SPH]. Sphinx4 is a flexible and modular framework. In addition to this it offers, as open source, a framework and several implementations examples. Based on these advantages we developed an approach based on the Sphinx4 technology. The Sphinx4 is composed of four entities: The 'Words in audio format' are words in digitized audio format, this entity is the input to the 'Voice recognition process' which is the core of the speech-to-text process. The 'Grammar' is a database composed of a dictionary with a classification and pronunciations of the words, a language model that identify the words dependencies and an acoustic model that identify the sound representation of each phoneme. The fourth entity is the 'Words in text format' which is the output result in textual format of the speech-to-text process.

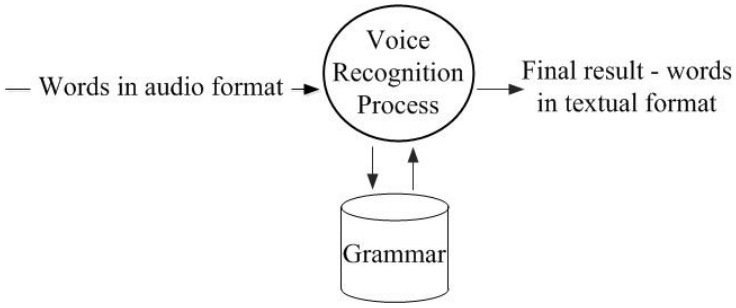


Figure 7.6: Speech recognition system - Sphinx4 process steps

The speech-to-text process is composed of four steps (Fig. 7.6): (1) The audio words to be recognized are sent to the voice recognition process, (2) the voice recognition process compares the content to the existing grammar, (3) the nearest interpretation of the audio in the grammar is then returned to the Voice recognition process. (4) The words resulting from the process is then listed with a timestamp relative to each word, this result is referred as time-based or isochronic metatada 7.7.

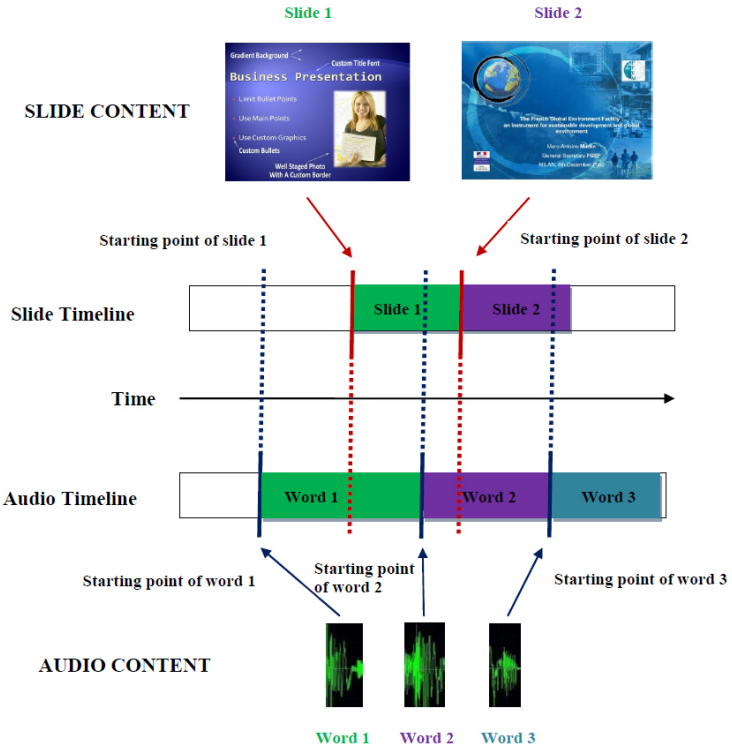


Figure 7.7: Speech recognition system - time-based metadata

In order to store the audio time-based metadata we used the MPEG7 tech-

nology. Indeed, the MPEG7 is a domain independent standard for multimedia description based on the XML language which allows us to offer interoperability with other systems. Its main goal is to provide structural and semantic description mechanisms for multimedia content. It provides content description for audiovisual content for describing different types of information namely lowlevel audiovisual features, highlevel semantic objects, content management and information about storage media. In addition to that, it offers possibilities for extension which is of interest for a domain specific tuning of the MPEG7 schema.

Improving the efficiency of the speech-to-text algorithm is an important point. Two optimization techniques were combined in order to enhance the result quality: On the one hand by removing the rare words, and on the other hand by modifying the segments durations of the audio track.

Algorithm concept:
The relevance value defines the importance of a word. This relevance can be compared to the page rank in Google for instance. When searching for a word in Google, a list of websites is displayed. This list starts with the results which are very relevant. Defining relevance in speech-to-text plug-in is not a trivial task. In fact, we only know the words which were spoken and the time when they were evoked within the video. Based on this information, a relevance value needs to be computed. To get a correct relevance value it is important to define some criteria. We based the analysis on the following criteria:

- The relevance of the word in the entirety of the audio track. To determine this relevance we take into consideration the number of occurrences of each word in the entirety of the audio track. If a word occurs frequently, it is then considered as a relevant word within the video.

- The relevance of the word in the context. To determine this relevance we take into consideration the number of occurrences of each word within an interval of time. If a word occurs several time it is considered a relevant word. This relevance can define whether the word is relevant in the context of the respective current subject in the audio track or not.

- The relevance value in the proximity. This parameter is determined based on the time that separates two occurrences of a same word. For instance, a word occurring twice in a short period of time is considered as a relevant word.

The content based indexing can be partially automated using features extraction techniques such as image or speech analysis. With the video segmentation and the speech-to-text analysis algorithms, a set of information is automatically extracted. In the former case, low-level features techniques were used in order to segment the video according to the slide-changes detection, in the latest a more high-level content information were automatically extracted in order to offer time-based metadata

information. The extracted information can be further enhanced by inputs provided by a manual and semi-manual content based annotation which would offer a richer and more semantic description of the video recordings such as information about the speaker, about the topics, the related projects, the location, and so on. The following section present the content based annotation approach based on the manual description guided by the conference knowledge model.

7.2.2 Manual and Semi-manual Content Based Annotation

Annotations are an important key-success in the emergence of the semantic web. The manual annotation may be a time consuming task. As stated by the state of the art, several interesting projects have been conducted in order to facilitate the manual annotation of multimedia. In this thesis work we focused on the manual and semi-manual annotation of videos and more precisely the videos of the recorded talks. We made use of existing tools and adapted them to conferences requirements by integrating on the one hand the video segmentation algorithm described above in order to offer the users the possibility to manually annotate, if desired, specific segments of a video recording, and on the other hand the knowledge model HELO in order to guide the users during their manual annotation process.

Every step of a conference life cycle requires some effort to generate information that can be reused. The information generated through the automatic extraction processes may be further enriched by manual annotation produced by users and also by semi-manual annotation provided by the aggregation of additional information coming from the conference life cycle (Fig. 7.8):

- **Administration Step**: The information at this level concerns the global information about the conference such as the location, the program of the conference, the conference main topics, the participants, the keynote speakers, etc. This information can be automatically integrated in the conference model. It mainly enriches the *TypeScope* (the category to which the conference is related), *TemporalScope* (the date/time when the conference takes place, the program of the conference and in some cases the related event such as the social event or co-organized conferenced), the *LocationScope* (the city, building, room where the conference takes place) and the *TopicScope* (the main topics defined for the conference).

- **Conference Multimedia Resources Acquisition**: The information at this level concerns the video recordings of the conferences talks and the related slide-show presentation. Once the video is generated, a link to the recording and the related presentation resources can be automatically integrated into the conference model. This level mainly enriches *MediaScope* (the media resources related to the conference).

- **Conference Related Multimedia Resources Acquisition**: The information at this level concerns the data related to the conferences talks such as the

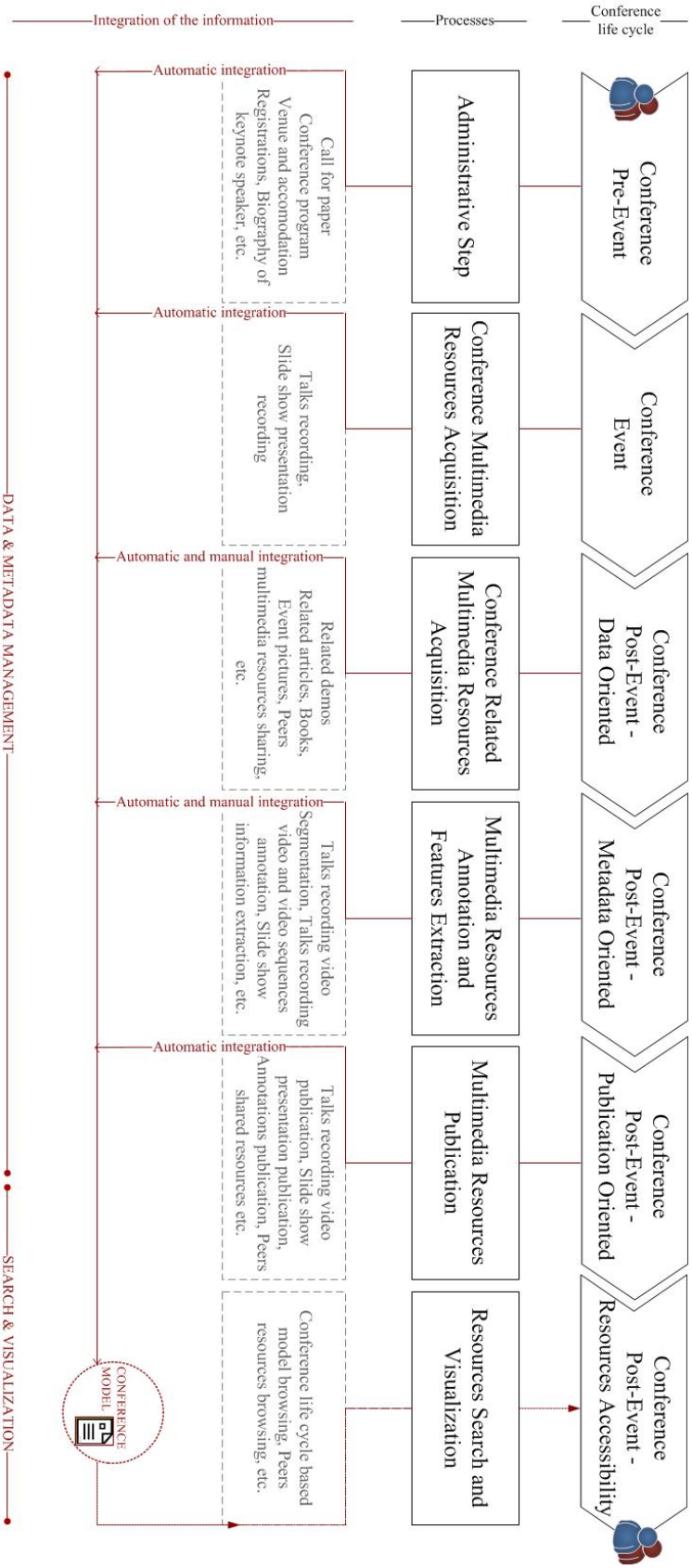


Figure 7.8: Conference life cycle information integration into HELO

papers, the books, pictures, websites, etc. This information can be manually added by the organizers of the conference but also by peers that would like to add their pictures, or related papers, etc. A link to the related data resources can be automatically integrated into the conference model. This level mainly enriches the *MediaScope* (the media resources related to the conference).

- **Multimedia Resources Annotation and Features Extraction**: The information at this level concerns the description of the existing resources mainly the description of the video recordings and specific sequences of the video recordings. This information is provided as stated above by automatic content based extraction process namely the video segmentation and speech analysis algorithms, and the peers manual model based annotations. This level mainly enriches the **TemporalScopeScope** (the video segmentation algorithm allows to automatically segment and assign to each video segment a start time and an end time, the speech analysis algorithm assign time-based keywords), the *ThematicScope* (peers annotations concerning thematics), the *EventScope* (peers annotations concerning the event itself or the related events), the *PersonScope* (automatic information extraction from existing resources on the web, peers annotation related to the persons) and the *CommunityScope* (peers annotation related to the community).

- **Multimedia Resources Publication**: This level concerns the available information (data and metadata) that has been generated along the previous step in addition to the information that can be found at peers level. The main purpose is to exploit the force of connected and interested peers which can bring further information. The link to the published information is integrated within the conference model. This level mainly enriches the *MediaScope* (the media resources related to the conference shared by connected peers).

7.3 Conclusion

Along the data and metadata management process the data resources can be automatically, manually or semi-manually annotated based on the conference model. In this thesis work we propose an hybrid information management that makes use of the different approaches (automatical, manual or semi-manual) in order to enrich the data resources description and therefore their retrieval. Indeed, in addition to the information provided by the automatic feature extraction algorithms for the video segmentation and the speech-to-text recognition, the information conveyed along the conference life cycle is captured in order to be used also within the annotation process. This information may be of different formats (MPEG7, RDF, OWL, etc.) which raises the issue of handling the retrieval process over heterogeneous metadata. The following chapter presents the knowledge driven retrieval approach.

Chapter 8

Knowledge Based Searching

8.1 Introduction

As described in the previous chapter, a conference life cycle conveys a set of information that may, if captured, enhance the multimedia information retrieval process. In this chapter we focus on the retrieval process. We present the proposed approach for the multimedia searching (querying and browsing) and visualization based on the knowledge model. The section 8.2 introduces the thesis early consideration related to the knowledge based querying approach. The section 8.3, presents the knowledge based multimedia browsing and visualization.

8.2 Multimedia Knowledge Querying

The coexistence of different data formats and consequently different query languages, leads to the need of handling the structural and the semantic heterogeneity of information. To handle this issue we designed a query model named Virtual-Q presented in this section. This model handles mainly three parts: (1) Dispatching important search information depending on the data sources, (2) creating database-specific queries and (3) merging results from several sources.

8.2.1 Virtual Query Engine

In order to provide users with a transparent and easy access to retrieve information from heterogeneous data sources, we propose an approach that integrates concepts such as the query analysis or the sub-queries formulation both of which facilitate the transparent access to heterogeneous data. Existing approaches that have been briefly presented in the chapter 3, rely often on elements that make easier the sub-queries reformulation, such as a global and a local ontology for each data source used only with the XML query (XQuery) techniques [XQU], or a human intervention (from an administrator for example) in order to determines which similarities and conflicts exist between the different sources and to define the rules for inter-schema correspondences. In this thesis work we propose an autonomous (data structure independent) query engine able to query over heterogeneous data sources by taking advantage of semantic and clustering techniques. The designed model, Virtual-Q, is illustrated in the figure 8.1. It is composed of four main parts: (1) The user interface, (2) the Virtual Query Engine (VQE), (3) the data and metadata storage (generated through the conference life cycle) and (4) the external reasoning models.

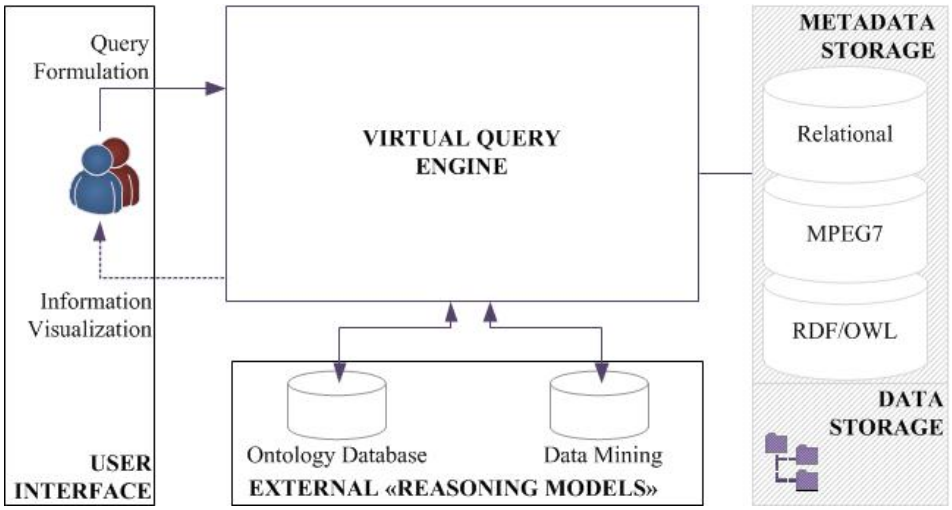


Figure 8.1: Virtual-Q - an overview of the approach

The autonomous query engine referred in this thesis work as the Virtual Query Engine (VQE) (Fig. 8.2), is composed of several modules. These modules are categorized into two parts: The query process modules and the resource management modules.

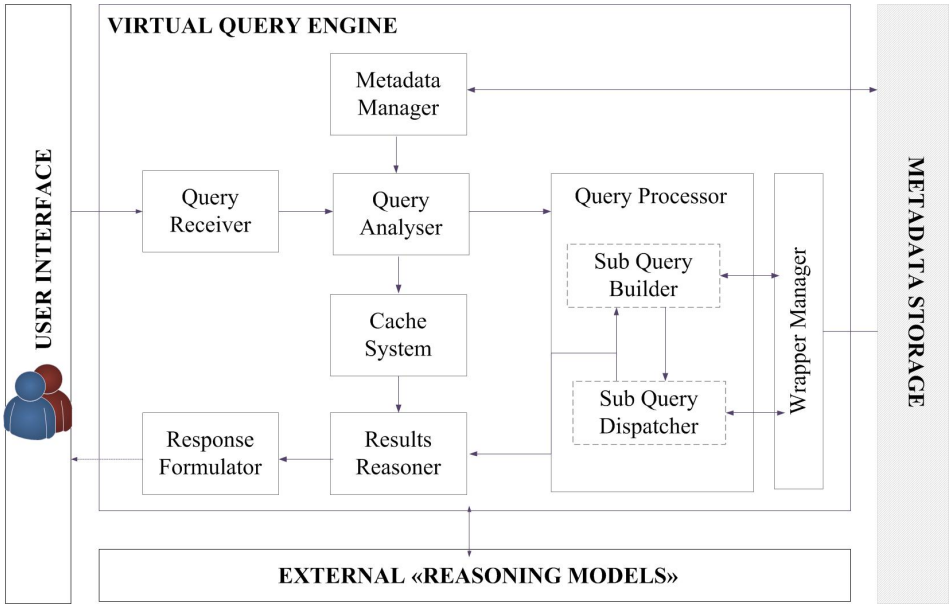


Figure 8.2: Virtual Query Engine (VQE) architecture

Query process modules:

- Query Receiver: Receives the original user query and transforms it in the format used inside the Virtual Query Engine.
- Query Analyzer: Analyses the user query in order to offer the users the pertinent data. It starts by retrieving the most adequate metadata schema helped by the Metadata Manager and the Wrapper Manager modules. Then, with different methods analyses the query and stores those data in the query. We talk about Virtual Query.
- Query Processor: Composed of the Sub Query Builder and the Sub Query Dispatcher modules. The Sub Query Builder generate a first query to be run on the most adequate data source. It transfers this query called virtual query to the Sub Query Dispatcher. The Sub Query Dispatcher send the virtual query to the concerned wrappers. Once the results are returned the Sub Query Dispatcher merges them and send them to the users through the Result Reasoner module.
- Results Reasoner: Reasons on the results guided by external *reasoning models* to return the results, delete redundant ones, etc.
- Response Formulator: Takes charge of compiling the queries results and prepare them to be sent to users in a adequate format.

Resource management modules:

- Metadata Manager: manages the metadata, connected data sources, installed wrappers, topics with source ranking and reasoning sources, etc.
- Wrapper Manager: is responsible of forwarding the queries to the concerned wrapper(s).

The following section presents the query process over the Virtual Query Approach.

8.2.2 Query Process

The query process is reported in the figure 8.3. The user expresses the query through the user interface. The query is then transmitted to the VQE. It passes through the chain of modules: (1) The user enters the query which is then transmitted to the engine. (2) The query is analyzed according to the existing metadata schema. (3) Based on this analysis, different queries (named virtual queries) are rebuilt to reach the different data sources format. When possible, the virtual query addresses the most adequate database namely the semantic database. Indeed semantic databases may enrich the original query, adding to it related keywords, and building additional virtual queries on top of the new discovered keywords. (4) After receiving the results, the VQE performs reasoning process to rend the results, delete redundant ones, etc. (5) The results are compiled and

formatted into the adequate format output. (6) Finally, the results are displayed to the user.

One of the most important issues of this query process is the query analysis and the virtual queries building steps described here below.

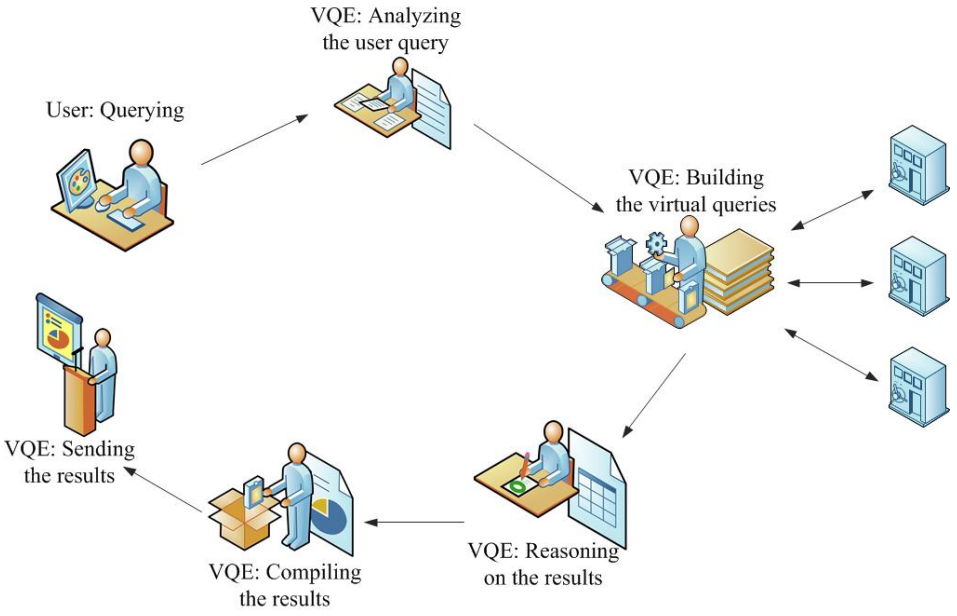


Figure 8.3: VQE process

8.2.2.1 Query Analysis

Since we want to make transparent the data access and minimize the participation of the user, the engine should integrate intelligent modules able to retrieve the most pertinent information. The query analysis is designed in this approach:

- Data sources schemes: Each data stored into a source is compliant to a schema or model. Depending on the type of the source, some schemes may be more complex and complete. In most cases, every source is automatically linked to its schema.

- Ranking the schemes: In some cases a specific schema may afford more information than another one. Namely, the ontology models provide more structured and rich information than a simple Google Desktop XML SchemaGoogle Desktop Search is a locally executed service that lets users index and search the contents of their computer. It provides a set of schemes for indexing files. The base schema, called Google.Desktop.Indexable stores information such as textual content, document length, format, etc.. At this level the VQE ranks the existing schemes according to the most complete regarding the requested query.

- Using a schema: Based on the result of the schemes ranking, the more adequate schema is exploited to refine the building of virtual queries (using associated metadata, inference features, etc.).

8.2.2.2 Virtual Queries Building

Based on the query analysis, the VQE builds the virtual queries (called also sub-queries). For instance, building a Structured Query Language (SQL) query for a relational database "*Select X From Y Where Z*" or an XQuery Full-text search for an XML document. The sub-queries are then dispatched to the wrappers through the Sub Query Dispatcher. The wrappers can be of 'Full text search' type, 'API-based' or 'fields-based reformulation'. In the API based wrappers, we consider the software sources that are accessed through an API. such Google Desktop or Windows Search Desktop. For this kind of wrappers, and according to the type of sources, we could directly forward the query. The developer of the wrapper has to transform the results returned by the API into the XML format to be merged later. For full text search, we are searching through the entire document (structure and data). Some APIs already exist for XML and could be used. If we consider searching into XML documents, the results do not need further changes before being merged. The field-based reformulation refers to the data sources where the structure is used to retrieve data. The free user query has to be transformed into a specific language using its associated syntax.

A summary of a query process is described through the figure 8.4-8.5. The chapter 10 presents early tests over the VQE implementation.

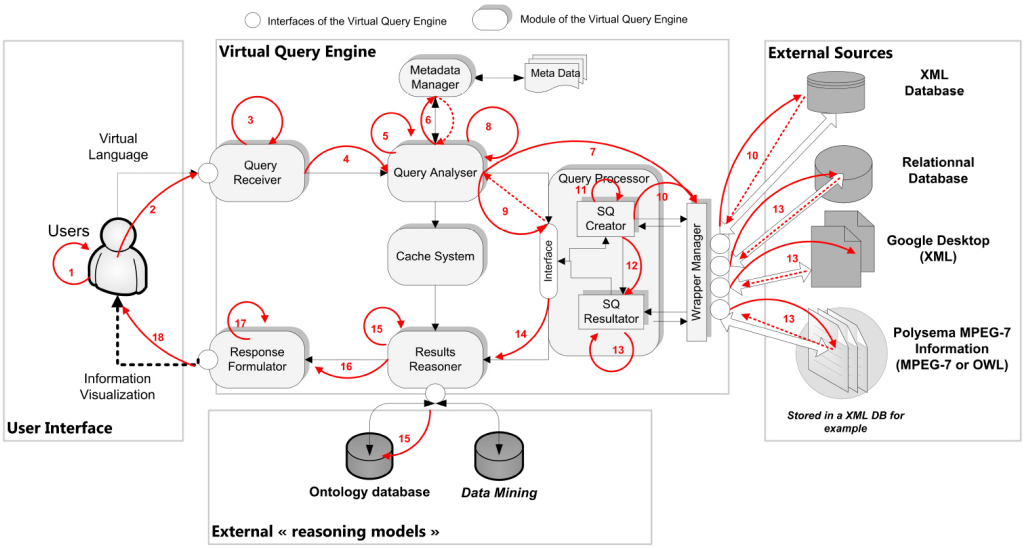


Figure 8.4: VQE - query process steps

	Description	In	Out
1.	The user writes his query using the graphical interface. The UI is not important, because we just have to specify the interface and if necessary define some helpful *relations* such theme=travel or limit=30. This information could be expressed by the user or secretly added by the application.	-	String
2.	The query is transmitted to the Virtual Query Engine as a simple string. The UI manager would be responsible to transcribe the query depending on the UI to a string.	String	AQ
3.	The first module transforms the general *AQuery* into *UserQuery* which contains the XML query. The query is kept in the XML document as a String element. The element also contains information about the source and the result format.	String	UQ
4.	The query is forwarded to the Analysis module.	-	-
5.	The analysis module transforms it into a *VirtualQuery* Element. Quickly done seeing that VirtualQuery is an extension of UserQuery.	UQ	VQ
6.	Using the metadata, the module checks the best schema that could be *downloaded*. In the metadata, we can find for each available theme, a ranking of the data sources about the pertinence of their schema. For example, an ontologies database has a better data model than the Google Desktop engine. A ranking could be Ontologies, XML, Relational, generic,..	Th	DS
7.	It requests the schema of the corresponding sources through the Wrapper Manager. This function must be implemented in the wrappers because it is specific to each data source. The wrappers chooses how maintains the schema.	DS	Sch: XML
8.	Using the schema, it analyses the query and creates a *StructuredQuery*. This can help us to find relations between the terms and for example provides some keywords.	VQ	VQ
9.	The Query Processor module receives the virtual query. It transmits it to the SubQuery Creator in its internal chain.	VQ	VQ
10.	It contacts the wrapper of the best data source according to the structured information that has been discovered using its schema. The wrapper creates and run the query. The results are returned to the module.	VQ	VQ
11.	It checks the obtained results. Those could be useful to refine the VirtualQuery regarding fields names for example.	VQ	VQ
12.	The VirtualQuery is forwarded to the SubQuery Resultator module.	VQ	VQ
13.	Next queries are run using the wrappers which build correct query according to the results or searching for a *mapping*. *Free queries* could be used if no matching are found.	VQ	VQ
14.	The SubQuery Resultator merges all the results and transmits them to the Result Reasoner module	VQ	VQ
15.	The Result Reasoner *reasons* on the information. Therefore, it could use additional external sources containing ontologies or data mining. It filters, sorts, groups...	VQ	VQ
16.	The VirtualQuery is forwarded to the final module,. the Response Formulator.	VQ	VQ
17.	The Response Formulator arranges the results as they must be. This is the output interface of the Virtual Query Engine.	VQ	XML
18.	The results are returns to the user interface that will display them.	-	XML

Figure 8.5: VQE - query process steps description

In this section we present early reflections for the integration of heterogeneous and distributed data sources. The proposed virtual query engine architecture attempts to offer a transparent access for users to retrieve data from heterogeneous data sources. In the rest of this work we focused rather on the browsing and the visualization issues than the querying issue briefly presented in this section. The following section presents the knowledge based browsing over the multimedia resources.

8.3 Multimedia Knowledge Browsing and Visualization

We presented in the chapter 6, the semantic model, which characterizes the knowledge conveyed within a scientific conference life cycle. In the chapter 7, we described what kind of information can be extracted and associated to the multimedia data. In this chapter we focus on the browsing process according to the knowledge model. We propose two retrieval modalities: A model based interface for performing a request and a faceted browsing interface to search into the resources. The model based interface allows the users to express complex request on top of the knowledge model in a simple way. The faceted search allows the users to explore the content through the semantic model concepts without the need of formulating a precise query. It is also a useful tool for the verification, and consequently the refinement, of the audiovisual content in a relatively simple, precise and interactive way. Browsing helps the users to discover and to learn both the structure and the content of the information being accessed. In this sense, browsing is also a means for understanding rich information contents. These features provided by browsing become essential for the retrieval of multimedia data resources, which have a complex content, not obvious to search and retrieve by using only querying mechanisms. The browsing is realized based on the underlying knowledge model, which defines the different navigation possibilities provided to search the content.

8.3.1 Model Based Interface

The knowledge model, as described in the chapter 6, was designed to allow on the one hand a granular description through the annotation oriented set and on the other hand to allow the formulation of complex request through information browsing based on a set of eight *scopes* (Fig. 6.2). In an environment such as the one of scientific conferences the users may be of different types namely professors, PhD Students, industrials, and the like. In this context, an adequate browsing interface is required which provides users with multiple possibilities for content browsing and exploration. An example of a complex request required in such an environment would be "*Find a talk sequence recording of the conference held in Fribourg, in 2007, in relation with the event RCSO. The speaker works in the MISG group and obtained the best paper award*". As this example shows the request should take into account different aspects related to the conference: *MediaScope* (talk

sequence recording), *LocationScope* (held in Fribourg), etc. A request involving the eight *scopes* may be "*Find a talk sequence recording of the workshop held in Fribourg, in 2007, in relation with the event RCSO project meeting where the speaker talked about the memoria-mea project and presented a demo. The speaker is Stefano Carrino's colleague and works in the MISG group and wrote the paper A survey of multimedia retrieval*" (Fig. 8.6).

Find a talk sequence recording	↔	*MediaScope*
of the workshop	↔	*TypeScope*
held in Fribourg,	↔	*LocationScope*
in 2007,	↔	*TemporalScope*
in relation with the event "RSCO project meeting"	↔	*EventScope*
where the speaker talked about the "memoria-mea" project	↔	*ThematicScope*
and presented a demo.	↔	*MediaScope*
The speaker is Stefano Carrino's colleague	↔	*PersonScope*
and works in the "MISG" group	↔	*CommunityScope*
and wrote the paper "A survey of multimedia retrieval."	↔	*MediaScope*

Figure 8.6: Example of a complex request in natural language

This request expressed with technical languages such as SPARQL (Fig. 8.7) is hardly accessible to every user. We offer users a conference model based interface (Fig. 8.8). Using this interface, the users have the possibility to execute traditional keyword based searching or perform more advanced searched based on the conference model through the different *scopes*, having the possibility to go even deeper in the granularity by searching into the different concepts defined in each of the eight *scopes* (Fig. 8.8). Using HELO as the underlaying semantic model is

```
PREFIX geo: <http://www.geonames.org/ontology#>
PREFIX helo: <http://www.semanticweb.org/ontologies/2009/4/HELOtheOrigin.owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?videosegment WHERE {
?talkevent helo:hasDate "2007"^^xsd:gYear.
?talkevent helo:hasRelatedEvent
<http://www.semanticweb.org/ontologies/2009/4/HELOtheOrigin.owl#RCSO_Projects_Meeting>
```

Figure 8.7: Example of a complex request SPARQL query language

a powerful key in conference multimedia information retrieval. Still, to be effective the rendering of the browsing process and the results has to be easy to read and simple to navigate through.

8.3.2 Faceted Browsing

As stated by the state of the art, several visualization techniques exist to represent the information in a more suitable form to the human cognitive processing system. The human factors play an important role, including ability to navigate, relate,

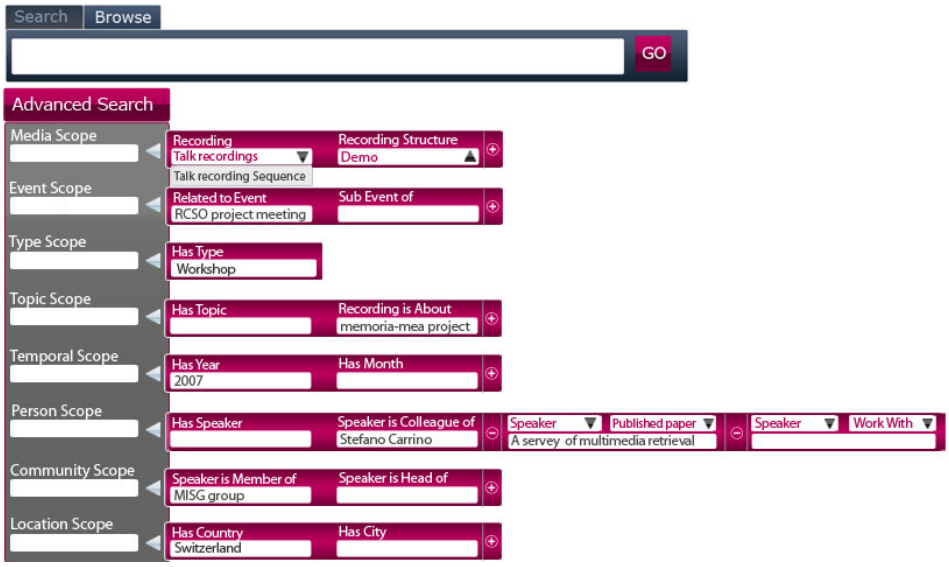


Figure 8.8: HELO based querying interface approach

remember and understand complex information, and possibly to collaboratively share and use knowledge. One of the approaches discussed in the state of the art is the faceted search. This approach combined to other visualization techniques used in this thesis work may have a significant potential to improve tasks related to knowledge retrieval by making it more accessible, understandable, sharable and dynamic.

The faceted approach allows the users to explore the available information which is very helpful if the user does not have a precise idea of the needed data. Using this approach the users may visualize the search result according to different *scopes* such as *PersonScope* (Fig. 8.9), *CommunityScope* (Fig. 8.10), *TemporalScope* (Fig. 8.11), etc. giving them a wide view of the scientific community network or a visualization through the *EventScope* giving the users a global view of the different events. The users may have a mix view by choosing to visualize the result according to more than one *scope* such as *PersonScope* and *TemporalScope* or *TemporalScope* and *LocationScope* (Fig. 8.12). This approach is essential and interesting due to the data filtering aspects which, depending on the cases, may further enhance the visualization of the search result.

When choosing a scope, the user has the possibility to zoom inside the scopes feature. For instance when choosing the *PersonScope* (Fig. 8.9) we visualize the result as a social graph offering a map of the connections between the person represented by a node interlinked by a relationship. By passing over an arrow linking two nodes the concerned nodes are highlighted and the type of relation is popped. By choosing a node we set the focus on the chosen node, the related multimedia is then provided (such as the set of video proposed within the nodes

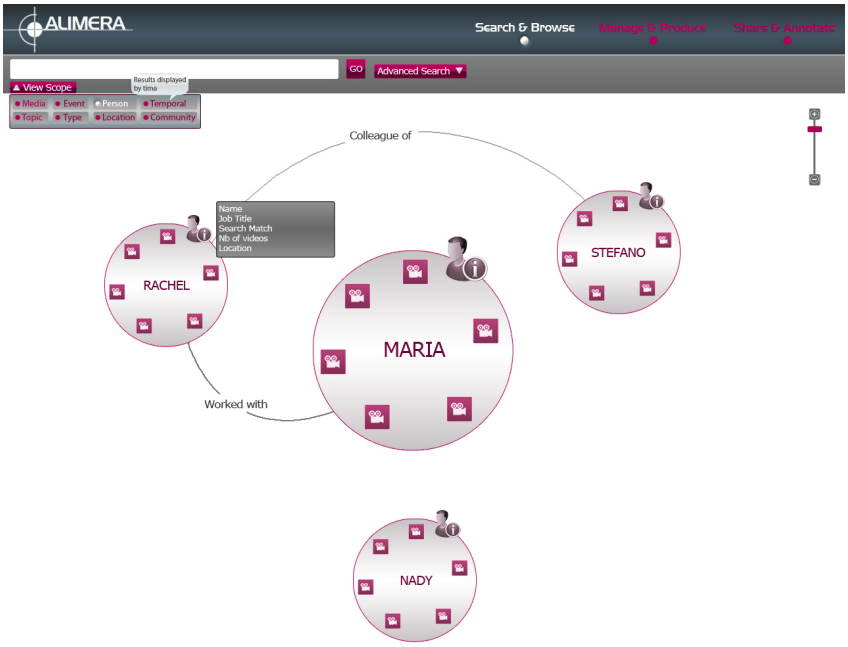


Figure 8.9: *PersonScope* interface approach

in the figure 8.9). This offers to the users the possibility to search within the multimedia resources related to the person selected and to filter these results according to any other *scope* such as the *TemporalScope* and *LocatioScope*.

Once the media is retrieved, the users have the possibility either to play the video within the search interface (an integrated interface offering a global view of the result, Fig. 8.13) or to get redirected to a new window where the video of the recorded talk is played within a conference based visualization interface. The visualization interface presents three distinct blocks: The recorded talk block, the slide set and the information block. The recorded talk is synchronized with the slides extracted from a presentation slide set. The navigation banner at the bottom of the slide set block allows the user to select any slide from the presentation in order to get the corresponding talk video sequence played. Using this interface to replay the recorded talks provide users with simple and easy navigation through the different sequences of the talk, in addition to the display of annotations, provided by HELO, for each video sequence. At any time the users can go back to the search interface to perform other queries.

8.4 Conclusion

Thanks to a combination of powerful existing techniques different aspects have been put forward in the proposed approach for a knowledge based searching in order

Figure 8.10: *CommunityScope* interface approach

Figure 8.11: *TemporalScope* interface approach

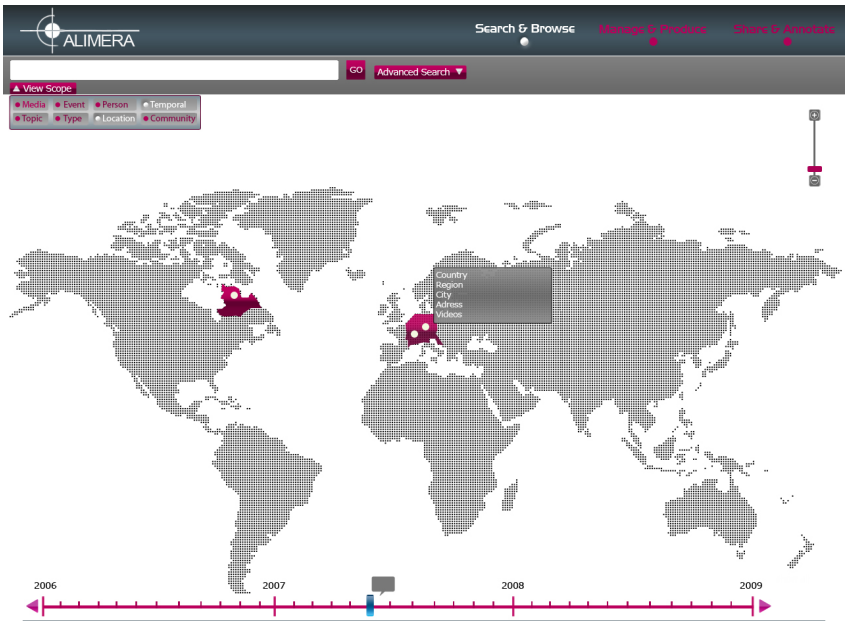


Figure 8.12: *Multi-Scope* interface approach

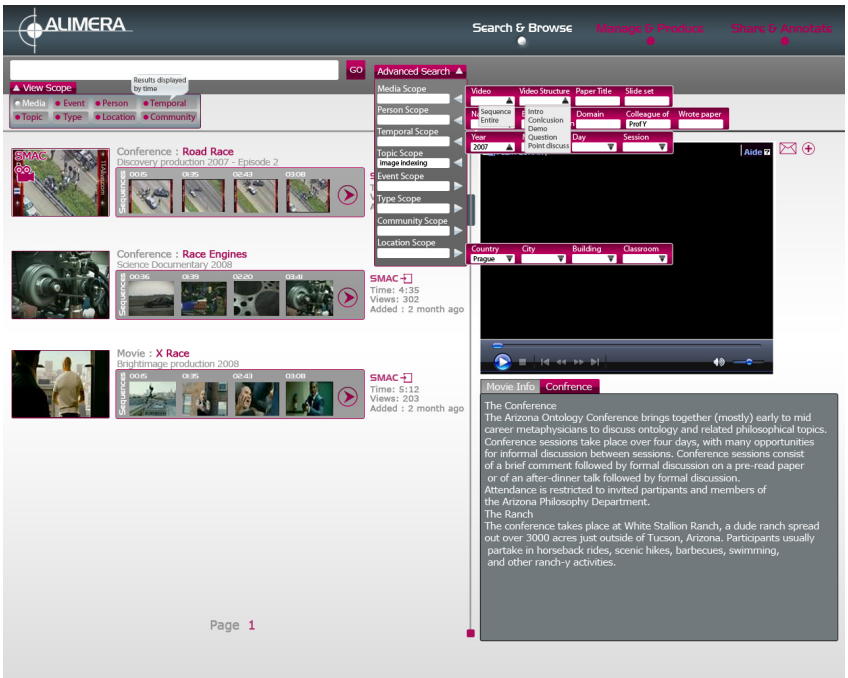


Figure 8.13: Results of the visualization interface approach

to enhance the conference multimedia retrieval. We addressed the issue of the knowledge based querying over heterogeneous data by proposing a Virtual Query Engine approach. Lacking of time we decided to provide this issue a minor focus. As stated earlier, the focus of this thesis goes mainly to the knowledge browsing and visualization comparing to the knowledge querying. We designed browsing and visualization interfaces based on the conference knowledge model which was designed based on the conference life cycle. The thesis proposed approach offers the users the possibility to perform complex requests and to visualize the results in a simple and interactive way. In the next chapter, we present the conference multimedia retrieval framework based on the model HELO and which addresses mainly the knowledge management, browsing and visualization issues.

Chapter 9

Conference Multimedia Framework

9.1 Introduction

The overall analysis of the state of the art led us to a set of observations and reflections which represent the main research axes of this thesis work. In the chapter 6, we presented the conference knowledge model that was designed according to the conference life cycle. This model was used along the conference knowledge management and in the knowledge browsing and visualization. Based on the observation done in the state of the art we propose in this chapter the framework CALIMERA, an end-to-end integrated framework for content and context based video retrieval. It is guided by the conference ontology model HELO. CALIMERA is composed of various designed approach for both knowledge based retrieval (management and searching) in the context of scientific conferences. In the section 9.2, we describe the framework approach. In the section 9.3, we present the use of the P2P network in order to enrich the multimedia retrieval process of CALIMERA.

9.2 CALIMERA Framework

The framework unifies the many efforts that have been conducted so far and go further by integrating novel approaches, based on the conference life cycle, within the multimedia information retrieval process, namely the annotation, the browsing and the visualization side. CALIMERA offers scientific conference management framework which allows the administration of the conference, the recording of the talks, the video segmentation based on a specific algorithm and other algorithm for features extraction, the semi-manual annotation guided by the conference model and the publication and retrieval of the multimedia resources related to the scientific events. CALIMERA is composed of two main units: The knowledge management, and the knowledge retrieval. The figure 9.1 outlines the global view of the framework which is composed of the following components:

- **Tools manager**: CALIMERA is a tool independent framework. It allows, through the tool manager, users (peers) to integrate or use their tools that may be involved in data or metadata management, search and visualization or both. CALIMERA offers a basic set of tools for the conference life cycle process. It is composed of INDICO, SMAC, CALISEMA and INVENIO which all are described in the chapter 10.

- **Data and metadata management**: The data and metadata management is composed of three main modules. The resources acquisition, the features extraction and the automatic and semi-manual annotation. The acquisition part is responsible of the integration of the administration part of a conference such as the conference program, the registration, the venue and accommodation, etc. and the integration of the data acquisition such as the talk recording acquisition and the related resources acquisition. For the feature extraction and the annotation module, CALIMERA manages through the tools manager the tools and modules involved in the multimedia semi-manual annotation and automatic features extraction such as the slide-change detection and the video sequence annotation.

- **Data and Metadata storage**: The data and metadata storage is responsible of the storing of existing data and metadata. This storage stores data of different formats such as mov, avi, pdf, odf, doc. It stores different type of metadata formats such as the MPEG7 which is one of the most widely used standard for multimedia description, and RDF and OWL which are a more semantically oriented standards for multimedia annotation that integrates high-level semantic description.

- **Search and visualization**: The search and visualization is composed of three main modules: The virtual querying system to handle the heterogeneity of annotation standards and formats, the browsing module and the visualization module. This unit, is responsible of querying the data and metadata storage in order to return the video and/or the set of video sequences and the related multimedia resources to the users.

- **Conference model: HELO**: The designed conference model which describes the information conveyed within a conference life cycle is integrated in the framework in order to guide the knowledge management and the knowledge retrieval. This model provides users with efficient and granular search facilities and allows the retrieval of videos of recorded talks within conferences, based on semantic criteria.

9.3 CALIMERA P2P Approach

The retrieval framework is a combination of management and searching approaches which can be used in an integrated way. In a scientific conference context, browsing and accessing the resources is an essential point for users. We focused on the browsing approach which highly rely on the conference life cycle acquired knowledge. The browsing modality allows the exploration of the database content in a more free style. To that end we designed different visualization interfaces based on the HELO model and made possible the synchronization between the different media resources. In addition to these approaches mainly axed towards

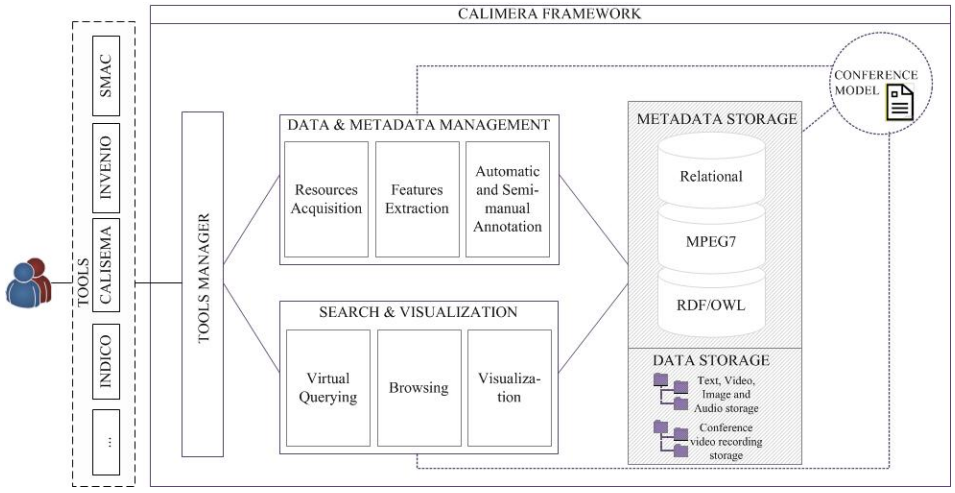


Figure 9.1: CALIMERA - global view architecture

the knowledge management and the end user requirement, we propose to take advantage of the potential of the P2P networks. The figure 9.2 presents the thesis proposed approach referred as the layered approach.

The proposed layered model maps the architecture of the CALIMERA. The layered model describes the adaptive approach offered by the framework aiming to guarantee an end-to-end service. The model is composed of four main layers, the Network layer, the Integrator layer, the Knowledge Management layer and the End User layer. Each layer uses performed actions of the underneath layer:

- **The Network layer** deals with the P2P protocols and network. Every peer of the Network layer may host a database to be shared with its peers in the network. Through the Network Manager of the Integrator layer every peer can access the P2P network. Each peer can make use of the Integrator layer to push information (through the Data Manager) into the knowledge database of the framework and can integrates knowledge management tools into the framework through the Tool Manager. The peers make use of the End User layer to perform tasks such as the information browsing and visualization.

- **The Integrator layer** manages the peer integration within the network on the one hand through the Network Manager and the communication between the peers and the knowledge database on the other hand through the Network Manger also. The Integrator layer is also responsible of the tools integration through the Tools Manager and the data integration through the Data Manager.

- **The Knowledge Management** layer handles the Data and Metadata Management (processing, indexing, querying, extracting, etc.). This layer is a key point for the End User top layer. It integrates, in addition to the Data and

Metadata Management, the Semantic Search Engine (composed of the virtual querying and browsing modules 9.1), and different modules for the Intelligent Data Processing such as clustering.

- **The End User layer** which focuses on the users needs and it is composed of the Model Based Data Annotation offering the user the possibility to annotate data guided by a conference model. It integrates also the modules involved for the Model Based Browsing and Information Visualization.

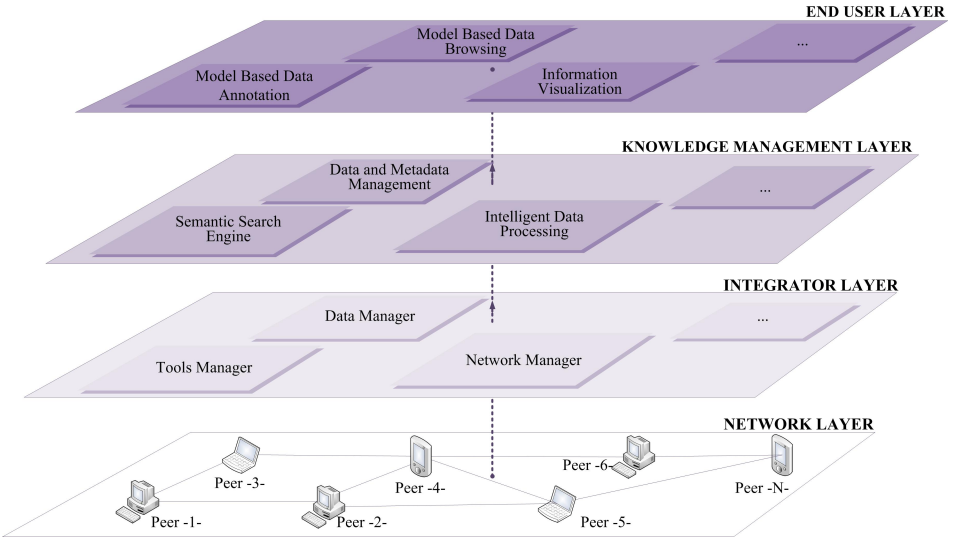


Figure 9.2: CALIMERA - layered model overview

9.4 Conclusion

In this chapter, we presented the CALIMERA framework and its layered model view. CALIMERA is a novel approach addressing the existing issues in video indexing and retrieval. CALIMERA is an integrated framework that aims at providing users with efficient and granular search facilities and allows the retrieval of videos of recorded talks of conferences, based on semantic criteria. CALIMERA is based on the ontological conference model HELO, a model that considers the entire information conveyed through a conference life cycle. We presented in this chapter the layered model that maps the architecture of the CALIMERA over a P2P network therefore taking advantage of the potential of the P2P. In the next chapter we present CALIMERA platform that has been implemented over a P2P network.

Part IV

PROTOTYPE - CALIMERA

Chapter 10

CALIMERA Platform and Testbed

10.1 Introduction

In this chapter we present the CALIMERA platform and its basic set of tools and modules designed for the entire conference life cycle management according to the semantic conference model HELO. The platform is implemented over a P2P network based on the JXTA framework. We aimed at validating the proposed experimental project over different a set of different types of scientific events.

In the following section we present the basic set of tools integrated in CALIMERA. The presentation of the tool has been categorized into the information management set of tools and the search and visualization tools. In the section 10.3, we describe the network implementation of CALIMERA. In the section 10.4 we present the testbed and the evaluation aspects of the platform. Finally in the section 10.5 we conclude the chapter.

10.2 Platform Tools

As stated earlier, the thesis proposed approach takes advantage of the content and the context information conveyed along a conference life cycle in order to bridge the "semantic gap". In this chapter we propose a set of tools for the data and metadata information management, and for the information search and visualization. All together, the integrated tools cover the conference life cycle (Pre-Event, Event and Post-Event). The figure 10.1 shows the different tools according to the conference life cycle. The section 10.2.1 presents information management tools composed of INDICO, SMAC (acquisition and features extraction modules), INVENIO (submission and features extraction modules), and CALISEMA. The section 10.2.2 presents the search and visualization tools composed of INVENIO (search module), NAVIR and SMAC (visualization module).

10.2.1 Management Tools

The management tools handle on one hand the data production and acquisition and on the other hand these tools manage the metadata production and features extraction. This section describes four different tools: The section 10.4.1.1 presents

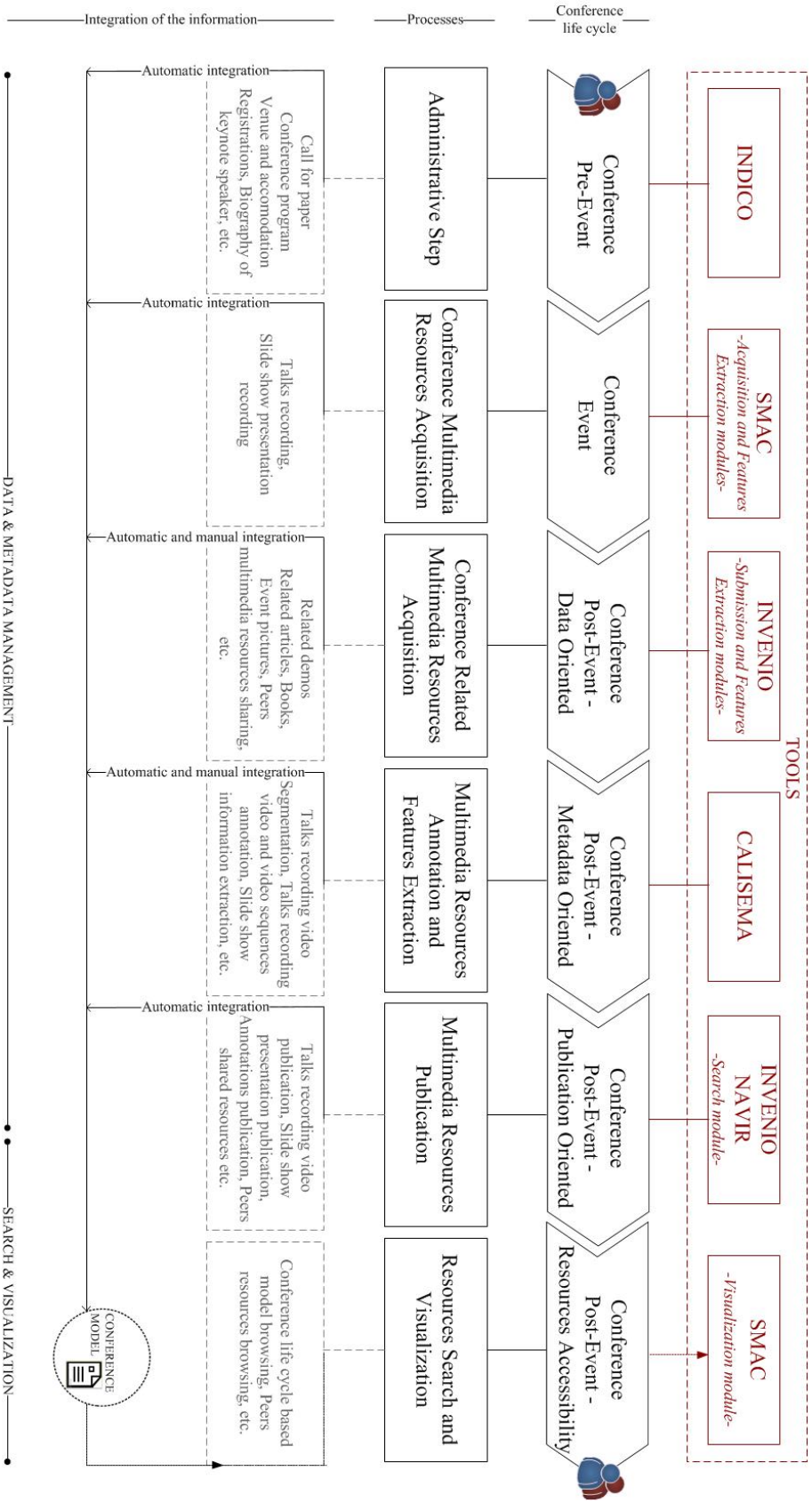


Figure 10.1: CALIMERA - conference life cycle set of tools

the administrative tool INDICO, the section 10.4.1.2 presents the tool SMAC with a focus on the acquisition and features extraction parts, the section 10.4.1.3 presents the tool INVENIO with a focus on the submission and the features extraction part, and finally the section 10.4.1.4 presents the annotation tool CALISEMA.

10.2.1.1 INDICO

INDICO was developed by the CERN to manage the administrative part of a conference such as the conference organization, planning, logistics and so on. We developed an additional module in INDICO in order to meet the thesis proposed approach. The adaptation concerns mainly the process. It was tailored in order to allow the inserted information to be integrated in the semantic model HELO and to automatically export the necessary information to the data acquisition tool which is SMAC. The figure 10.2 shows the different steps to perform in INDICO: The top interface in the figure 10.2 allow the users to initiate a new contribution, the middle interface shows the different fields to fill and the bottom interface allows user to export the necessary information to the data acquisition tool.

10.2.1.2 SMAC - *Acquisition and Features Extraction*

SMAC is a tool that we have developed which aim at recording the conference talks. It offers an automatic slide-change based video recording segmentation. This segmentation is performed according to the specific algorithm described earlier. SMAC is composed of different modules. (1) The Controller Module: It is the main module of SMAC since it is the reference point access for external applications namely INDICO. It controls the processes of the different modules. (2) The Recorder Module: It is responsible for acquiring the video streams through the Microsoft DirectShow application programming interface (API) which is a media-streaming architecture for Microsoft Windows. Using DirectShow allow us capturing the video and audio streams [DIR]. The Recorder Module is also responsible to acquire other type of information, mainly basic information such as the name of the speaker the topic and the like. The figure 10.3 shows the acquisition interface which automatically integrates existing information coming from INDICO and offer the possibility to add additional information and resources such as the presentation file. (3) The Analyzer Module: This module analyzes the captured video and the other digital documents provided to the system in order to temporally synchronize them. The current synchronization is done between the talk video recording and the slide-show presentation file (supported formats are MS-PowerPoint, OpenDocument presentations and PDF). The following section presents the video segmentation results using the proposed algorithm. (4) The Archiver Module: This module stores the streams captured during the conference (audio and video streams), the slide-show presentation file, the documents provided by the speaker and the output information brought from the analyzer module. (5) The Publisher Module: This module is responsible for creating an ergonomic user

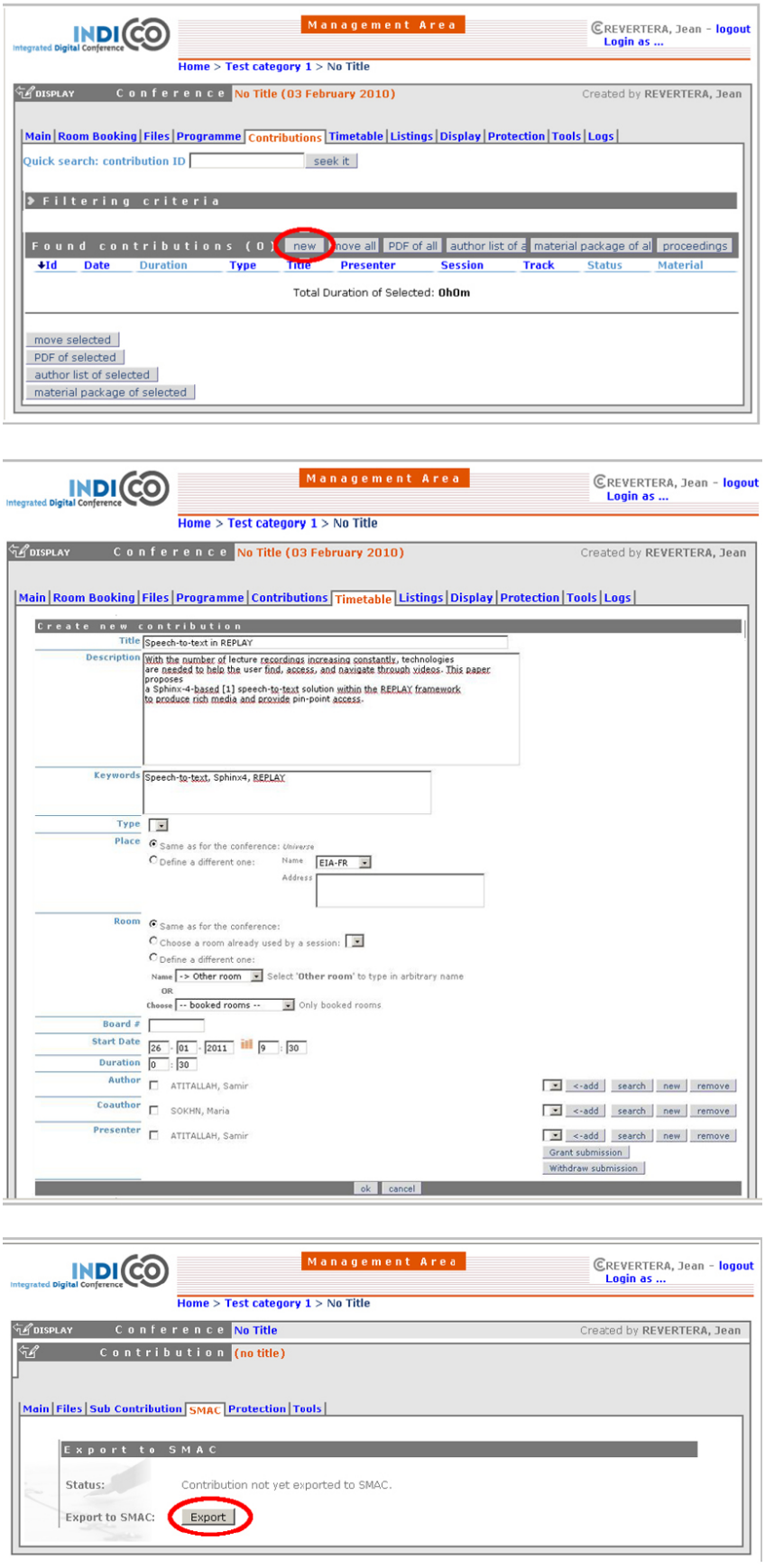


Figure 10.2: INDICO - conference management tool

interface for a talk video recording replay. This interface is described in the section 10.2.2.3. The major part of these modules is written in the Python language [PYT]. The acquisition software for accessing the capture device are implemented in C/C++ (via DirectShow). As shown in the figure 10.3 the generated information can be automatically exported to INVENIO for publication.

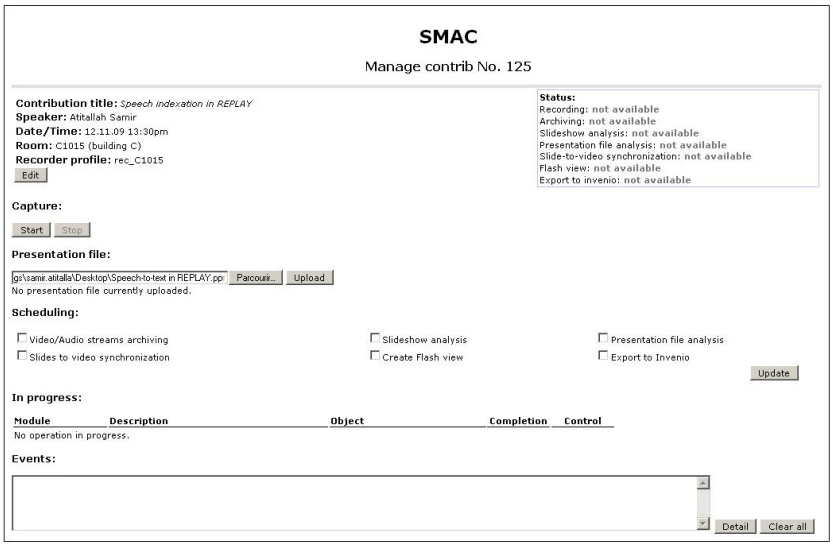


Figure 10.3: SMAC - acquisition interface

Video Segmentation System:
SMAC captures the slideshow of the speaker (Fig. 10.4) and records the talk. It

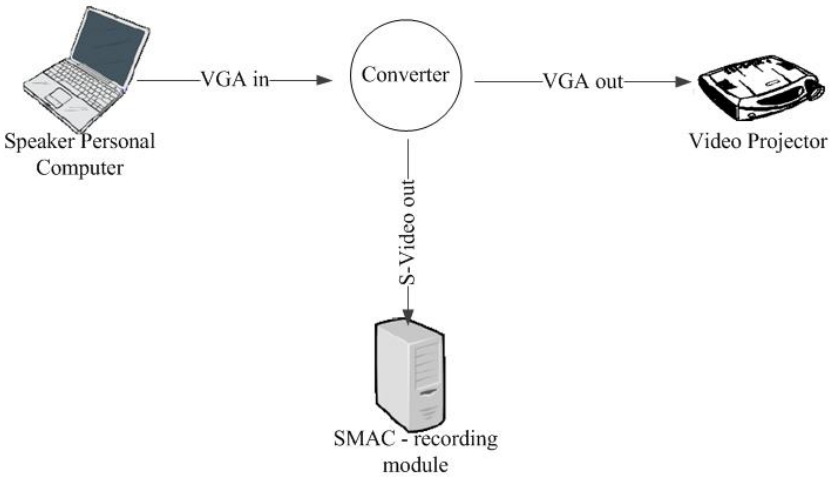


Figure 10.4: SMAC - slide-show capturing

automatically synchronizes the slides to the corresponding video segment. This process is performed through the slide-change based video segmentation algorithm.

Video	S_c	Q_n	M_o	Err_o
1	33	148	148	6
2	29	112	112	4
{3-60}	–	–	–	{0}

S_c: Number of effectively occurred slide-changes. Q_n: Total number of video detected sequence. M_o: Number of video sequences to slide matching. Err_o: Number of erroneous matching in M_o.

Table 10.1: Video segmentation system - sequence matching results

As described earlier the proposed algorithm approach is guided by the extracted information from the slides presentation and by heuristic hypothesis.

We performed the algorithm over sixty videos recording talks which included animated slides, external demos and non linear slides navigation. Over the sixty videos only two had a false sequence matching result. The table 10.1 presents the result. S_c is the effective number of slide-changes occurrence (during a talk, the speaker may go back and forward within his/her slides or even skip some). Q_n are the sequences detected by the video analysis. M_o are the number of video sequences matched to their corresponding slide using the proposed algorithm (obviously several sequences can be assigned to a single slide). Err_o, presents the number of falsely matched sequences. The obtained results prove the effectiveness of the proposed algorithm compared to previous works [Thomas M. 2007, John K. 1997, B. Erol 2003].

The video segmentation system was tested with a detection rate of one frame per second of video. The system performed with no false-negative on slide-changes, and with a filtering of most of the non-usable false-positive (Q_f value in the result table 10.2). The measured time of the execution of a one hour long presentation, is about sixteen minutes for 3600 frame comparisons. This means that the time of the execution is about 0.26 the real duration of the presentation. This result may be enhanced if we consider using the P2P potential.

The table 10.2 presents the final results of the proposed algorithm performed over twenty representative presentations with a considerable use of animations on the slides. The results show that the only use of identification approach (without subsequent refinement approach) is not sufficient enough for videos having less than 50% of their frames that may graphically correspond to a given extracted slide image. This proves the importance of the proposed refinement approach. The matching refinement takes no more than one second to be applied, we have noticed an average of 929ms. Table 10.2 description:

- T_i: Time during which the video displays a slide in an identifiable state, ie. when the displayed content corresponds to a slide picture extracted from the

Video	T_i	T_{nia}	T_{nio}	S_n	S_c	Q_f	Q_n $(Q_s/Q_a/Q_e)$	G_{id}	M_{no}	E_{no}	M_o	E_o
1	5′03″	3′55″	0′	12	14	0	41 (14/27/0)	25	23	0	18	0
2	18′05″	3′58″	0′	15	17	0	27 (17/10/0)	21	22	0	5	0
3	19′27″	0′	0′	19	18	0	18 (18/0/0)	18	18	0	0	0
4	29′08″	0′	0′	23	25	0	25 (25/0/0)	25	25	0	0	0
5	34′51″	0′	0′	20	20	0	20 (20/0/0)	20	20	0	0	0
6	15′45″	1′34″	0′	15	24	0	28 (24/4/0)	28	28	0	0	0
7	18′50″	3′58″	0′	21	23	0	43 (23/20/0)	30	35	0	8	0
8	22′26″	0′30″	0′	24	24	0	26 (24/2/0)	24	24	0	2	0
9	35′30″	0′	1′17″	17	48	17	59 (48/0/11)	48	28	0	31	0
10	17′41″	0′10″	0′	24	27	0	29 (27/2/0)	28	28	0	1	0
11	11′11″	3′34″	3′23″	15	19	255	103 (19/52/32)	19	35	0	68	6
12	21′12″	2′57″	0′	52	56	204	78 (56/22/0)	78	78	0	0	0
13	3′08″	8′02″	5′17″	32	33	347	148 (33/75/40)	59	88	0	60	0
14	11′47″	7′36″	16′09″	12	14	264	32 (14/5/13)	18	16	0	16	0
15	5′43″	9′08″	2′48″	11	12	215	93 (12/43/38)	27	26	0	67	0
16	13′25″	0′55″	4′21″	10	10	261	41 (10/8/23)	10	39	0	2	0
17	12′02″	2′32″	2′19″	19	15	168	88 (15/29/44)	26	23	0	65	0
18	8′12″	16′22″	0′	29	29	135	112 (29/83/0)	56	52	0	60	4
19	11′26″	3′22″	5′03″	26	42	322	152 (42/16/94)	51	38	0	114	0
20	16′05″	0′	7′07″	19	19	427	73 (19/0/54)	19	73	0	0	0

Table 10.2: Video segmentation system - global results and statistics

presentation file.

- T_{nia}: Time during which a displayed slide on the video does not match the picture extracted from the presentation file (e.g. because of animations), potentially causing the identification to fail.
- T_{nio}: Time during which the video displays a non-slide content (e.g. during off-slide demonstrations).
- S_n: Number of slides in the presentation.
- S_c: Number of slide-changes which effectively occurred. It could be $> S_n$ (at least one slide was displayed several time because of a backward iteration in the slide stream) or $< S_n$ (at least one slide was not displayed at all).
- Q_f: Number of sequences filtered by the passive mode.
- Q_n: Total number of detected sequence. It is composed from:
  - Q_s: Number of sequences related to slide-changes.
  - Q_a: Number of sequences related to in-slide activities (animations).
  - Q_e: Number of sequences related to off-slide activities.
- G_{id}: Number of sequences correctly identified.
- M_{no}: Number of sequences matched to a specific slide without orphan sequences assignment.
- E_{no}: Number of falsely matched sequences in M_{no}.
- M_o: Number of sequences matched through orphan sequences assignment.
- E_o: Number of falsely matched sequences in M_o.
- T_{syn}: Time taken by the synchronization process (in milliseconds).

10.2.1.3 INVENIO - *Submission and Features Extraction*

INVENIO is a free software suite co-developed by an international collaboration comprising institutes such as CERN, DESY, EPFL, FNAL, SLAC. It offers the possibility to put in place a document repository. The technology offered by the software covers all aspects of digital library management from document ingestion through classification, indexing, and dissemination. INVENIO complies with standards such as the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic format. INVENIO is being used by about thirty scientific institutions worldwide. INVENIO consists of several more or less independent modules with precisely defined functionality. The figure 10.5 presents the different modules that compose INVENIO. The details of each modules are described on the INVENIO website [INV]. We integrated into

INVENIO the speech-to-text plugin which is related to the BibClassify process. BibClassify allows automatic extraction of keywords from fulltext documents, based on the frequency of specific terms, taken from a controlled vocabulary. Controlled vocabularies can be expressed as simple text thesauri or as structured, RDF-compliant, taxonomies, to allow a semantic classification. In order to integrate the speech-to-text plugin we setup the configuration file in order to find the mplayer and the ffMpeg applications that are responsible of the separation of the audio track from the video flow. We also setup the video directory where the plugin should find the video. The speech-to-text plugin is a java application that is executed through a batch that runs periodically over videos that still does not have an MPEG7 output file.

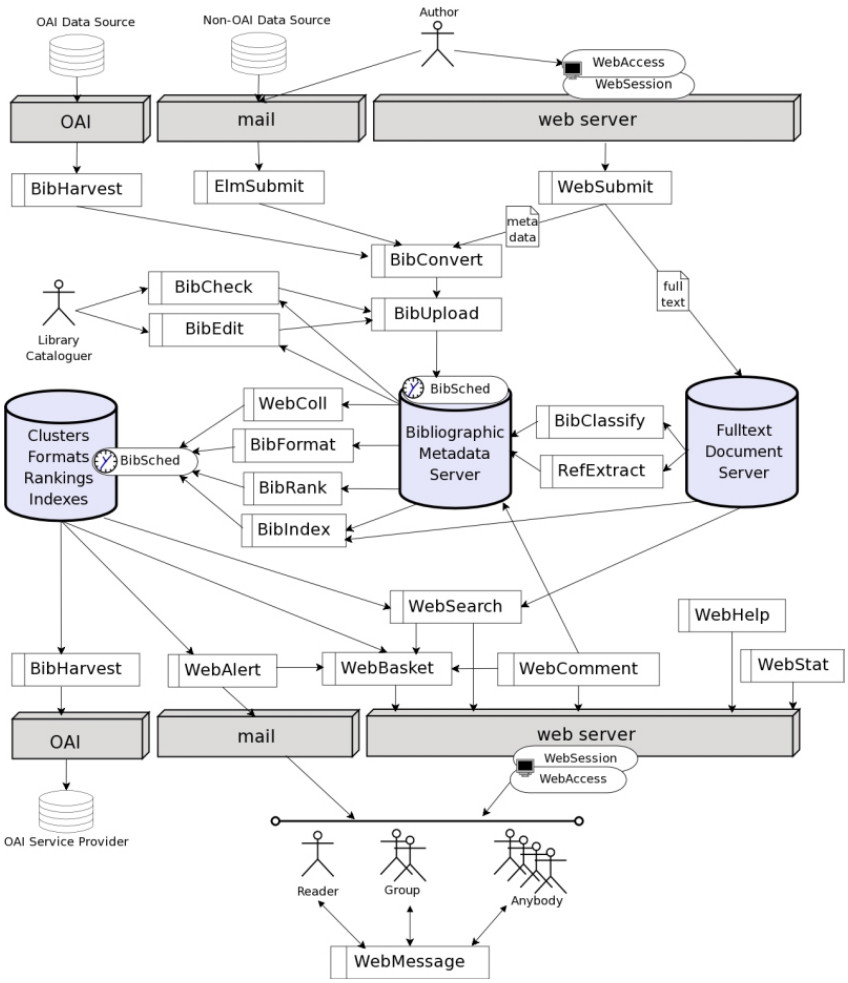


Figure 10.5: INVENIO - modules overview

Speech-to-text System:
As described in the knowledge management approach chapter, the proposed algorithm is based on three criteria: (1) The relevance of the word in the entirety

of the audio track, (2) The relevance of the word in the context and (3) The relevance value in the proximity. The implemented plug-in generates an MPEG7 file which describes all the audio content of the video (Fig. 10.6). This file provides the number of the segment (AudioVisualSegment), the starting time of the word (MediaRelTimePoint), the duration of the word (MediaDuration), the value between 0 and 1 which defines the certainty of the correctness of the audio recognition (Confidence), the value between 0 and 1 which defines the importance of the word in the audio track (Relevance), the language of the word (Lang), and finally the word itself (Keyword).

```
<AudioVisualSegment id="track-1.segment-49">
        <MediaTime>
        <MediaRelTimePoint>T00:00:36:899F1000</MediaRelTimePoint>
      <MediaDuration>PT90N1000F</MediaDuration>
    </MediaTime>
    <TextAnnotation confidence="1.0" relevance="0.6" xml:lang="en">
                <KeywordAnnotation>
                        <Keyword>security</Keyword>
                </KeywordAnnotation>
        </TextAnnotation>
</AudioVisualSegment>

<AudioVisualSegment id="track-1.segment-50">
        <MediaTime>
                <MediaRelTimePoint>T00:00:37:825F1000</MediaRelTimePoint>
                <MediaDuration>PT130N1000F</MediaDuration>
        </MediaTime>
        <TextAnnotation confidence="0.7" relevance="0.3" xml:lang="en">
                <KeywordAnnotation>
                        <Keyword>home</Keyword>
                </KeywordAnnotation>
        </TextAnnotation>
</AudioVisualSegment>
```

Figure 10.6: Speech-to-text algorithm - example of generated MPEG7 file

The tests have been carried over a set of videos which were recorded in different contexts. An example of the results is showed in the figures 10.7 and 10.8.
As shown in the figure 10.7, the improvements are significant. This is done thanks to the optimizations performed over the audio track which contains some specific words with weak pronunciation. However, this improvement costs information lost. The figure 10.8 compares the percentage of existing words in the video before and after the optimization. We notice a decrease in the percentage of the extracted words. This decrease is due to the optimization by removing the rare words. This approach has a drawback on relatively short video, since the correct words will not occur enough and may be removed by the optimization algorithm. In this case the proposed optimization should not be applied rather traditional speech to text integration may be applied.

10.2.1.4 CALISEMA

CALISEMA is a metadata management tool. It has been integrated into CALIMERA in order to facilitate the annotation process. CALISEMA has been developed in collaboration with the university of Athena. It has been adapted to

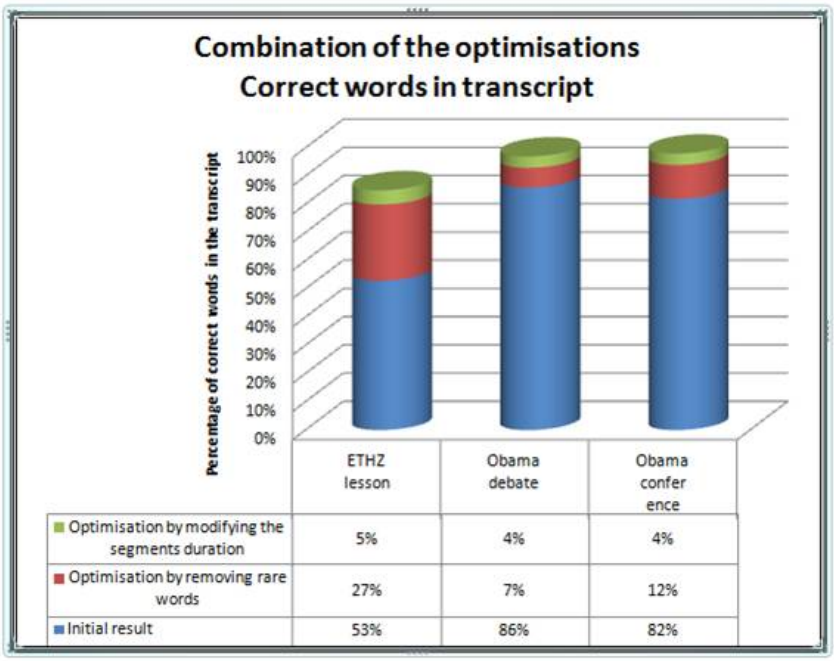


Figure 10.7: Speech-to-text algorithm - results of correct words in the transcript after the optimization

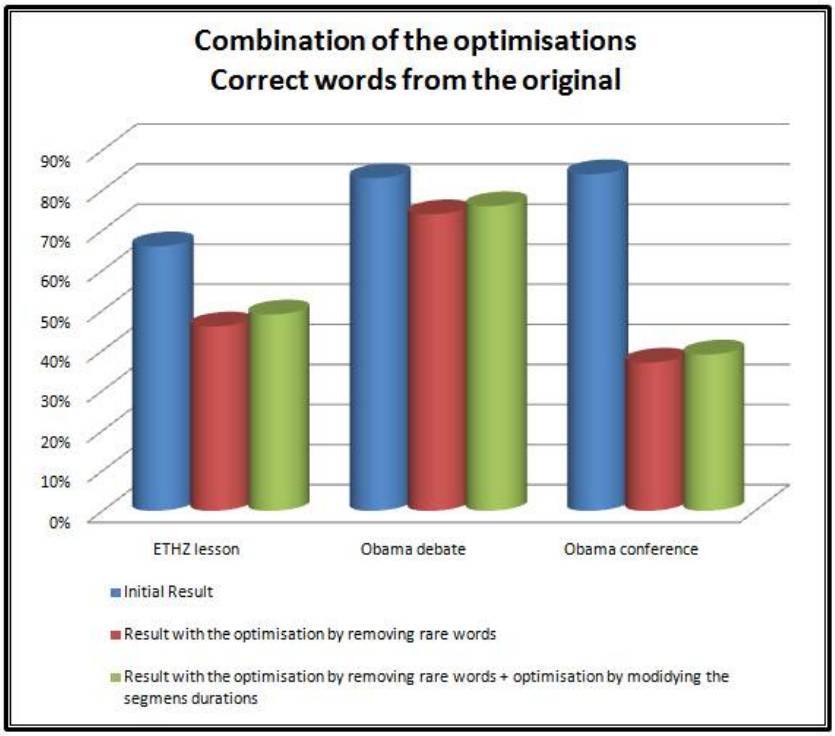


Figure 10.8: Speech-to-text algorithm - results of correct words from the original video after the optimization

integrate the CALIMERA platform. The figure 10.9 describes the main interface of CALISEMA. The bottom part -1- shows the keyframes of the video recording of the talk. Each keyframe corresponds to a slide in the talk presentation. The left part -2- shows the video sequence corresponding to the chosen slide. Each sequence can be annotated in several ways (top-right part -3-), such as an annotation guided by an ontology (bottom-right part -4-). CALISEMA has an algorithm

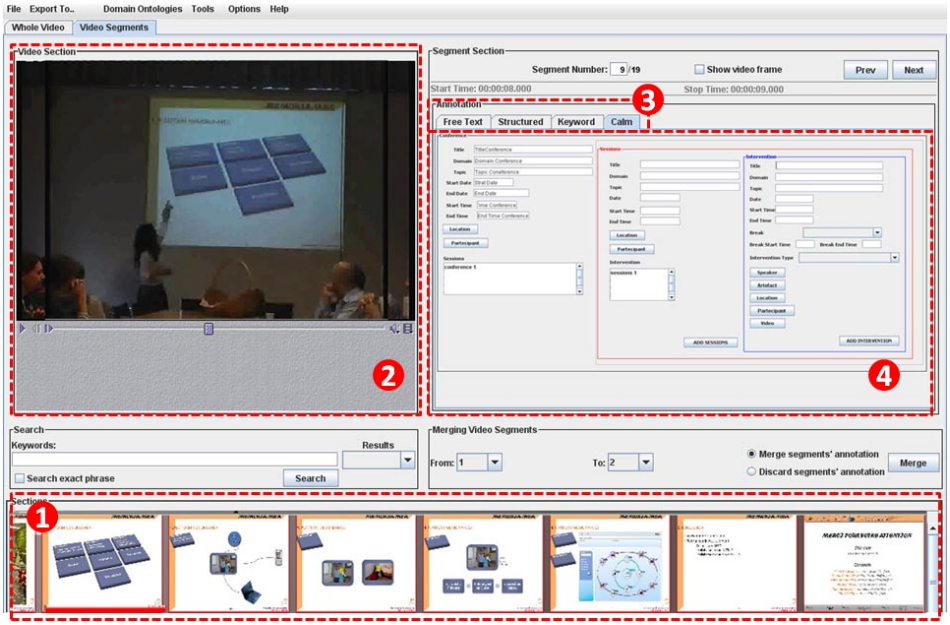


Figure 10.9: CALISEMA - conference model based video annotation

manager that allow users to choose the segmentation algorithm they want to apply on the video to be annotated through a graphical user interface (Fig. 10.10). By default it integrates an algorithm that segments periodically the video (e.g. every five minutes), an algorithms that segment the video according to the video scene changes, and the thesis proposed algorithm which segments the video according to the slide-change detection. We believe that this algorithm is the most suitable for conference recordings as it allows to segment a video according to the slide-changes and consequently to the talk. The figure 10.11, presents the CALISEMA interface, more particularly the customized dialog for the algorithm choice. By choosing the smac.jar algorithm, the user has the possibility to choose from the SMAC archiver module the talk video recording and its associated metadata namely the administrative information provided by INDICO and the video segmentation information provided by the SMAC analyzer module. CALISEMA has also a sequence manager that helps user handling (deleting, adding, or merging) the existing video segments. Each video segment that is represented by a keyframe, or a group of segments can be annotated through the CALISEMA interface guided by the HELO ontology. The figures 10.12 and 10.13 presents the annotation interfaces of CALISEMA driven by HELO. We can notice that some fields are already field thanks to the information

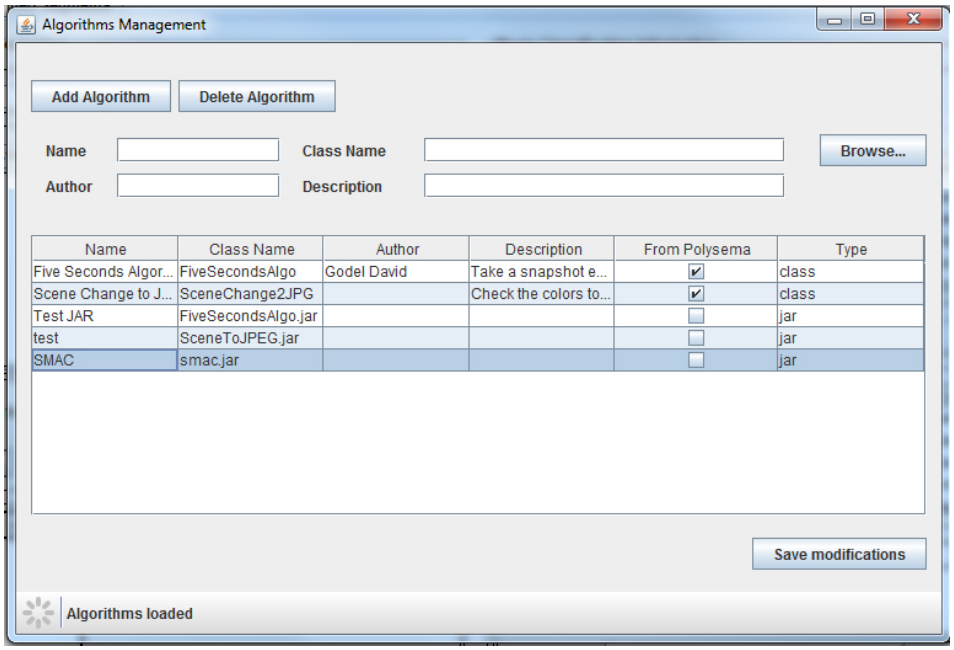


Figure 10.10: CALISEMA - algorithm manager

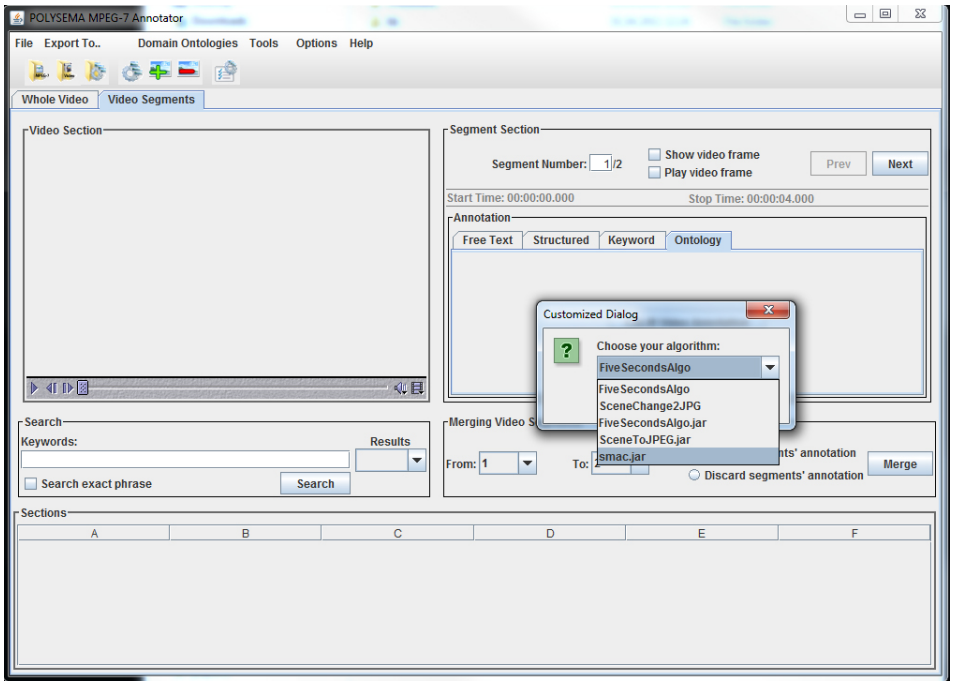


Figure 10.11: CALISEMA - slide-change based video segmentation use

captured from the previous process of the life cycle of the recorded talk. The description file is exported afterward in both MPEG7 and OWL formats which are pushed in the metadata repository.

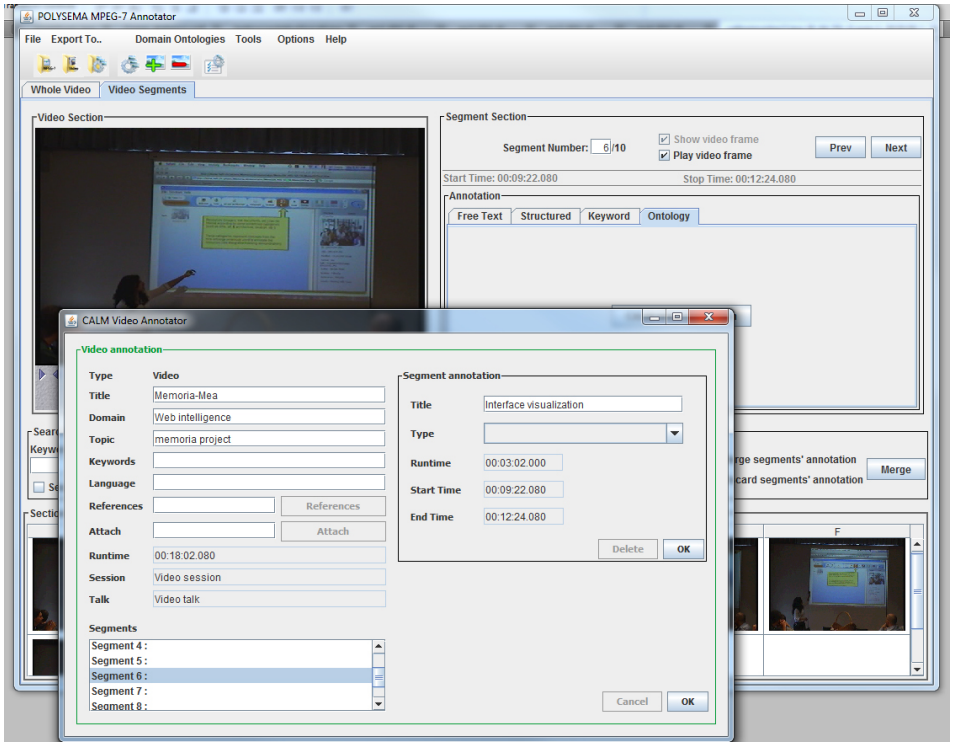


Figure 10.12: CALISEMA - automatically integrated information

10.2.2 Search and Visualization Tools

The search and visualization tools handle the publication of the different resources such as the talk video recordings, the related papers, pictures, events and so on. The section 10.2.2.1 describes the search module of the tool INVENIO and the early result of the Virtual Query Engine. The section 10.2.2.2 describes NAVIR, a tool ontology driven multimedia browsing and visualization, and finally the section 10.2.2.3 presents the talk recording visualization tool provided by SMAC.

10.2.2.1 INVENIO - *Search*

As described before, INVENIO offers the possibility to put in place a document repository, and the technology offered by the software covers aspects such as the dissemination. In this thesis work we tailored some of the modules in order to be able to automatically integrate, into the repository, the resources and the information generated along the life cycle of the conference more precisely information coming from INDICO, SMAC and CALISEMA. The figure 10.14 shows the traditional search interface into which we integrated a link to the talk visualization interface

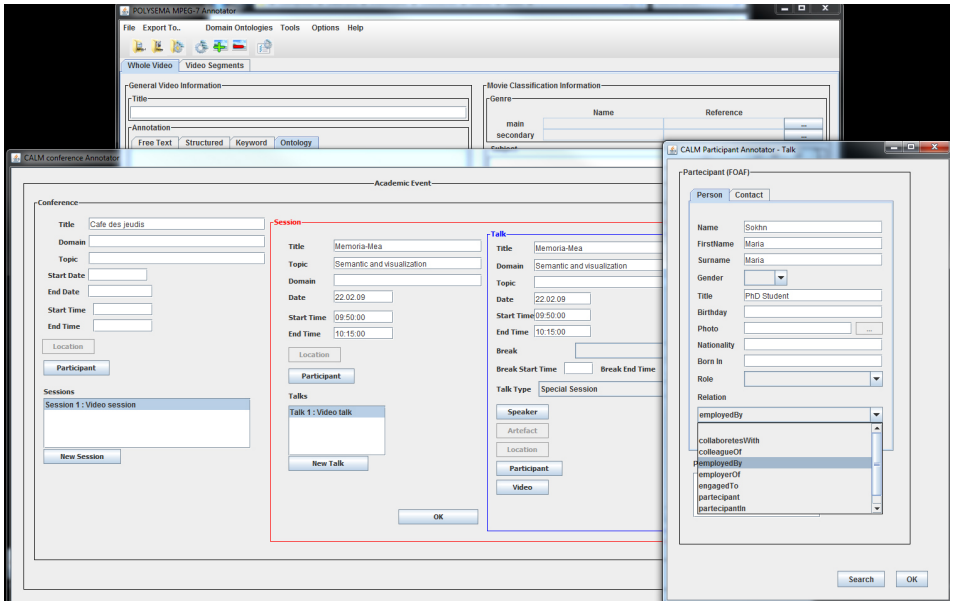


Figure 10.13: CALISEMA - speaker information integration

provided by SMAC. Each row is composed of basic information such as the name of the author, the title of the submitted article.

We have chosen to integrate, for the demonstration, the files created by CALISEMA and the ones created by Google Desktop as external sources. This choice was made for a twofold reasons: the first is to be able to query information generated by the annotation tools and the second is to query information stored on the user personal computer which is interesting in a P2P network. Two corresponding wrappers were integrated: PolysemaWrapper and GoogleDesktopWrapper. The PolysemaWrapper does a full-text search in the MPEG-7 files. It returns the files where the information has been found. the GoogleDesktopWrapper for its part connects to the GoogleDesktop search engine. It uses the URL of GoogleDesktop to run a query and retrieves the results in XML format. The XML schema provided by Google is the source schema for this wrapper. An example of the VQE is presented in the figure 10.15. The graphical interface is mainly made up of the administration window (Fig. 10.16, upper interface). It shows the connected data sources, the installed wrappers and proposes a log panel with the events. Two buttons on the tool bar allows addition of a data source or a wrapper. The graphical interface manages the data sources and wrappers. A query frame allows users to submit their queries (Fig. 10.16, lower interface).

10.2.2.2 NAVIR

NAVIR exploits the semantic model HELO. In addition to the traditional search tool, we propose an ontology driven searching interface mainly based on browsing

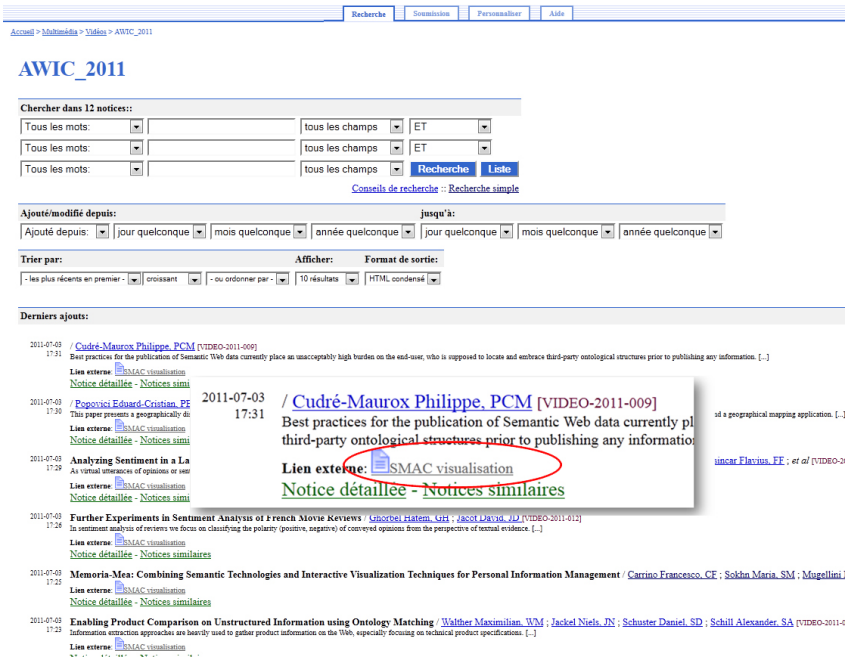


Figure 10.14: INVENIO - traditional search tool

activities. As described earlier, HELO is composed of a set of eight *scopes*: *PersonScope*, *LocationScope*, *TemporalScope*, *EventScope*, *ThematicScope*, *TypeScope*, *CommunityScope* and *MediaScope*. In order to validate the proposed approach, we implemented the *PersonScope* and the *ThematicScope* interfaces and we offer the users the possibility to filter their results according to one of the *TemporalScope* aspects which is the year. We have prioritized these scopes according to the following motivations: Regarding the *ThematicScope*, the *PersonScope* presents rich information, the person's information was more easily available in the proposed system (due to the mandatory fields about person in the conference management tool) and finally the current trends of social networks. Regarding the *ThematicScope*, Information about the thematic was also easily available (due to the different module of features extraction: Speech-to-text in INVENIO, slides analysis in SMAC), considering the reuse of one of the existing thematic classifications such as the ACM taxonomy [ACM].

PersonScope Aspects:

The *PersonScope* focuses on the scientific community members relationships and the multimedia data related to these people such as scientific publications and recorded talks. Among the existing libraries providing interesting navigation features, we can cite Processing [PRO a], javascript InfoVis Toolkit [INF], Raphael [RAP] and Prefuse library [PRE] available in a version for Adobe Flash technology called Flare [FLA b]. The Flare library was chosen to implement the prototype. Indeed, this library offers a wide range of different layouts such as the radial tree

```
ï»¿<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<VirtualQuery>
  <Information>
    <Person>
      <Name/>
      <Location/>
    </Person>
    <theme>Travel</theme>
  </Information>
  <OriginalQuery>japon</OriginalQuery>
  <DataSources>
 <DataSource best="true"><id>ds1</id><Properties>Polysema</Properties>< ↩
    Wrapper>w1</Wrapper></DataSource>
 <DataSource best="false"><id>ds2</id><Properties>Google Desktop</ ↩
    Properties><Wrapper>w2</Wrapper></DataSource>
  </DataSources>
  <StructuredQuery/>
  <Results>
    <file>
      <path>d:\Polysema\japan_graveyard.mp7.xml</path>
    </file>
    <file>
      <path>d:\Polysema\Japon.mp7.xml</path>
    </file>
    <results count="46">
      <result>
        <category>file</category>
        <doc_id>68281</doc_id>
        <event_id>74346</event_id>
        <title>test_complet.mp7.xml.mp7.xml</title>
        <url>D:\Polysema\test_complet.mp7.xml.mp7.xml</url>
        <flags>2</flags>
        <time>128648423053530000</time>
        <snippet>Title xml:lang=en type=main phoneticAlphabet=sampa ↩
            Super video du &lt;b&gt;japon&lt;/b&gt;&amp;lt;Title&amp;gt; ↩
             Abstract&amp;gt; FreeTextAnnotation phoneticAlphabet=sampa ↩
            Free text annotatation super super jadore dior</snippet>
        <icon>/file.gif</icon>
        <cache_url>http://127.0.0.1:4664/redir?url=http%3A%2F%2F127%2E0 ↩
            %2E0%2E1%3A4664%2Fcache%3Fevent%5Fid%3D74346%26schema%5Fid%3 ↩
            D4%26q%3Djapon%26s%3D7KBeL4IkmRwMsId3GBsmTTvzlnk&amp;src=1& ↩
            amp;schema=4&amp;s=NRpW3lFwVAT3F_1uSZQOVTr7o_4</cache_url>
      </result>
     // ... \\
      <result>
        <category>web</category>
        <doc_id>67052</doc_id>
        <event_id>72960</event_id>
        <title>godeld Bookmarks on Delicious</title>
  <url>http://delicious.com/godeld</url>
  <flags>259</flags>
  <time>128642967499050000</time>
  <snippet>firefox:bookmarks 63 Lien2 60 Cours 50 Projectdiploma 48 ↩
      WindowsXP 37 News 30 &lt;b&gt;Japon&lt;/b&gt; 28 All Tags 261 ↩
      Whats new? delicious about blog terms of service privacy</snippet>
        <thumbnail>/thumbnail?id=6%5Fvgk22D8dMQUAAAAA&amp;s=n4kZwSZW0- ↩
            R7KW7fdrn4i89BUs8</thumbnail>
        <icon>/icon?id=http%3A%2F%2Fdelicious%2Ecom%2F&amp;s= ↩
            JEniWW0e4DjEf6uvZsVl4AqdYCg</icon>
        <cache_url>http://127.0.0.1:4664/redir?url=http%3A%2F%2F127%2E0 ↩
            %2E0%2E1%3A4664%2Fcache%3Fevent%5Fid%3D72960%26schema%5Fid%3 ↩
            D2%26q%3Djapon%26s%3DjCHZYvlqaDlthAaZgtILNQ7aGyI&amp;src=1& ↩
            amp;schema=2&amp;s=yElWlQzM1mKb0cEPl6hjXuBQIco</cache_url>
      </result>
    </results>
  </Results>
</VirtualQuery>
```

Figure 10.15: VQE - output example

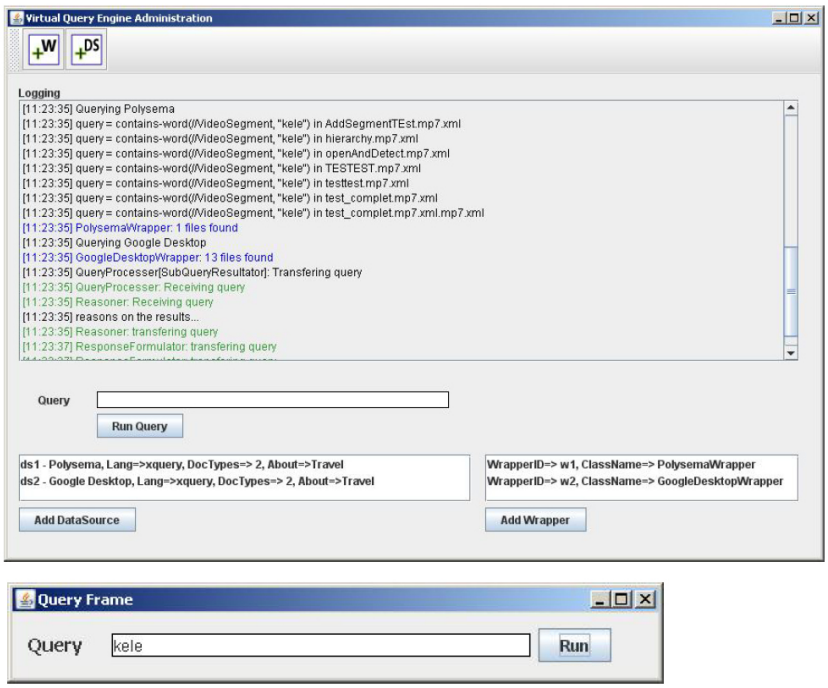


Figure 10.16: VQE - Administration interface and Query frame

layout or the circlepack (Fig. 10.17) layout which are needed for the prototype. It is an ActionScript library for creating visualizations that run in the Adobe Flash Player. From basic charts and graphs to complex interactive graphics, the toolkit supports data management, visual encoding, animation, and interaction techniques. It also allow users through its modular design to create customized visualization techniques without having to reinvent the wheel. Flare's design was adapted from its predecessor prefuse, a visualization toolkit for Java. Adding to this, Flare is an open source software.

The *PersonScope* interface aims at enhancing the browsing by offering a map of the connected people. The interface integrates several features facilitating the browsing and navigation. Each node represents a person and the links between the nodes represent the relationships between the people (Fig 10.18):

- Relationships: By passing the mouse over an arrow linking two nodes the concerned nodes are highlighted and the type of relation is popped.
- Focusing: By double clicking on a node we set the focus on the chosen node.
- Dragging: The nodes can be dragged, allowing to modify the nodes position.
- GraphDistanceFilter: The depth of the tree can change allowing a better overview of the social network related to the focused node.

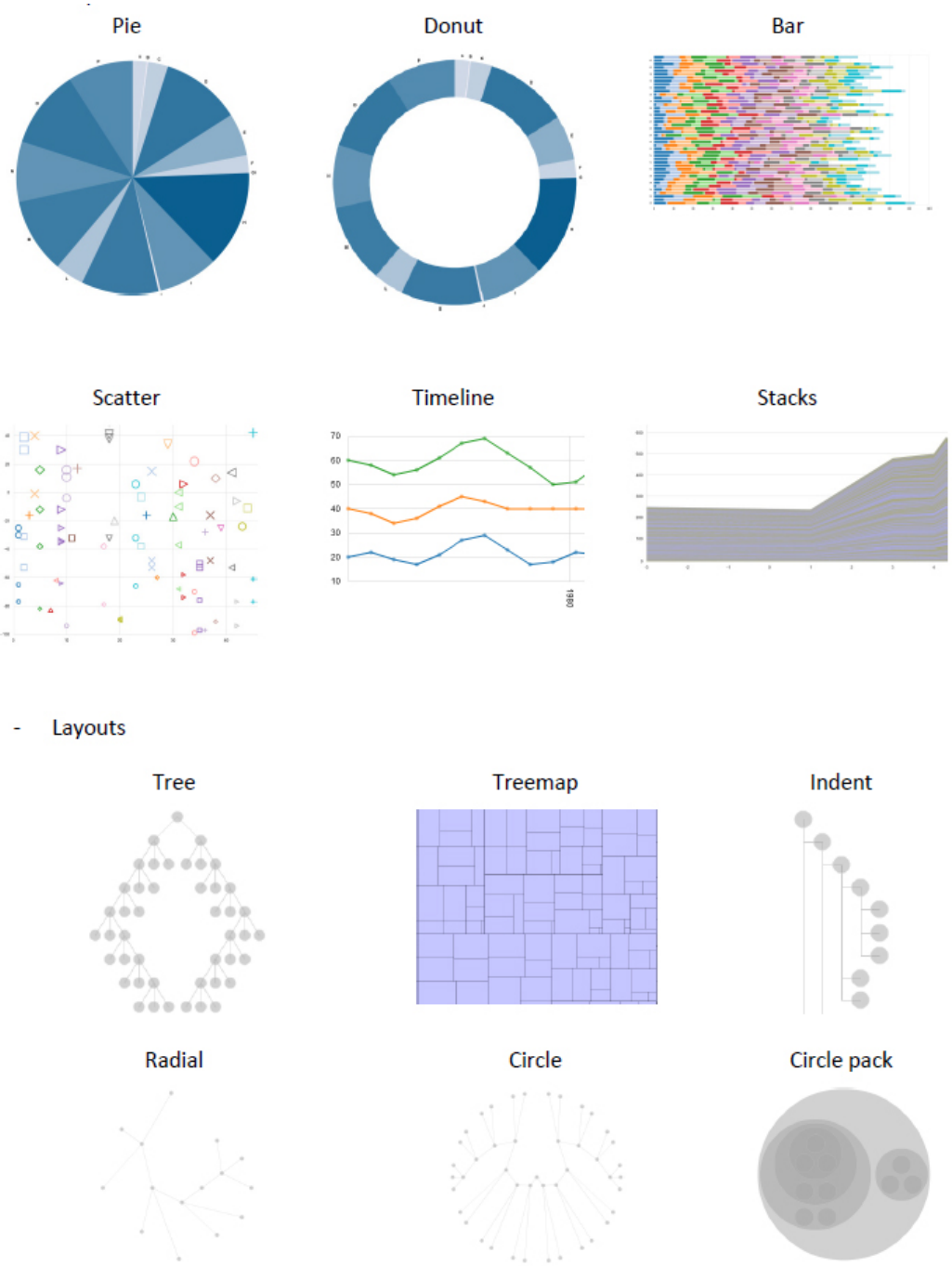


Figure 10.17: Flare - layout samples

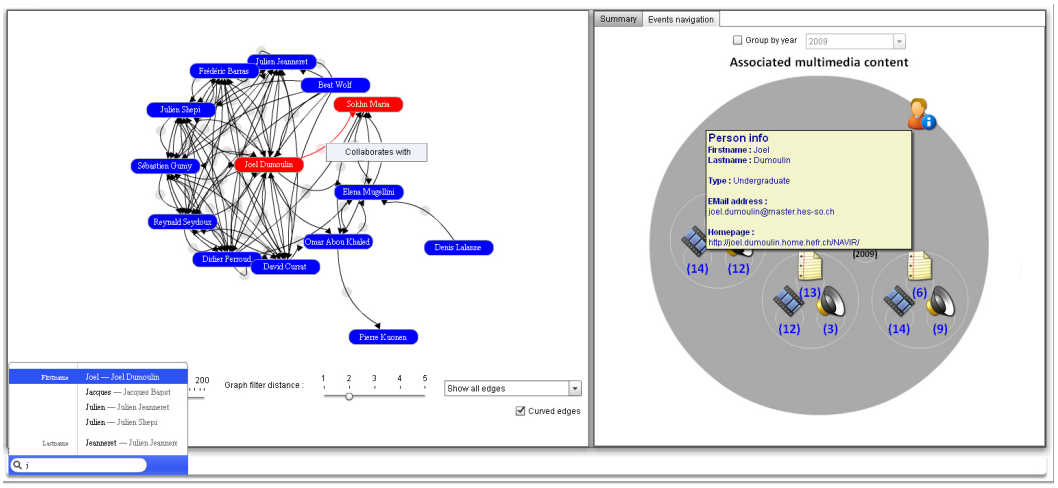


Figure 10.18: NAVIR - *PersonScope* implemented interface

ThematicScope Aspects:

The *ThematicScope* focuses on the different topics the recorded talks may have. It is based on the ACM taxonomy classification. We can browse according to the proposed topics keywords and if needed the interface offers the users the possibility to browse according to numbering convention of ACM. The results are grouped by type of media (video, audio and text).

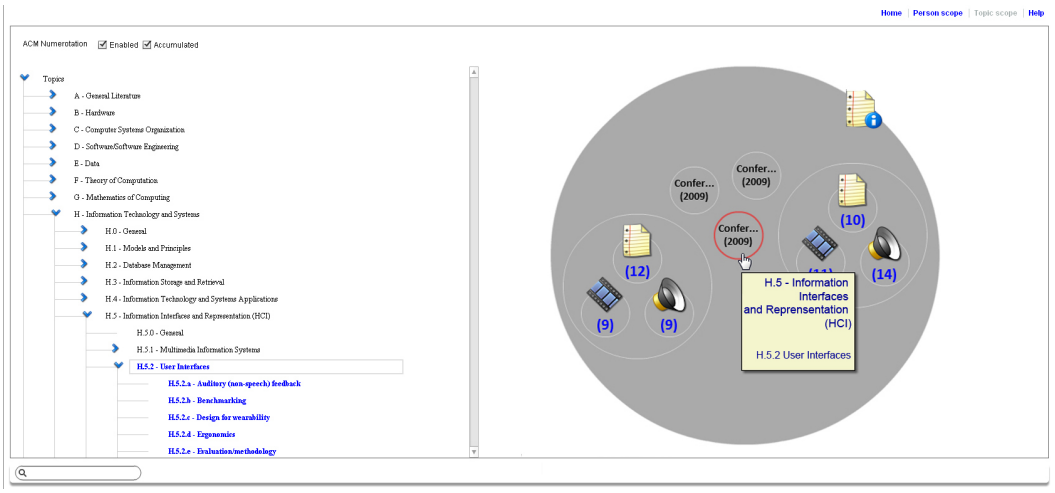


Figure 10.19: NAVIR - *ThematicScope* implemented interface

Multi-Scopes Aspects:

The *Multi-Scopes* or faceted filtering approach aims at offering the users a richer interface where several related information are available. It offers the possibility to

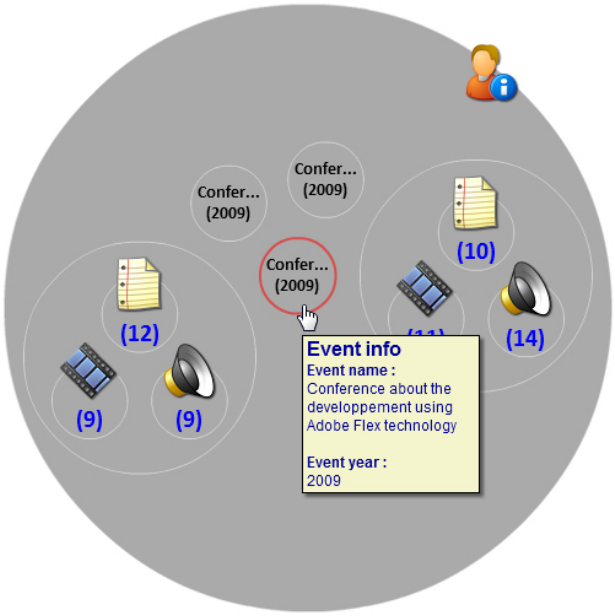


Figure 10.20: NAVIR - *EventScope* implemented interface

combine different *scopes*. The figure 10.18 presented above shows this approach: At the right side of this interface, the user can navigate within the multimedia related to the selected person displayed in the social graph. The displayed resources are grouped by *EventScope* and can be filtered by *TemporalScope*. Several features are offered by this interface:

- Person info: By passing the mouse over person icon, general information relative to the selected person is displayed (Fig.e 10.18, right part).
- Event info: By passing the mouse over an event circle, information about the event is displayed (Fig. 10.20).
- Related multimedia resources: By clicking on an event we expend the content composed of the related multimedia resources namely the recording talk, the audio resources and the textual resources.

10.2.2.3 SMAC - *Visualization*

Whether the users come across a talk recording through the INVENIO or NAVIR interfaces, they access to the SMAC visualization interface which is composed of three distinct blocks (Fig. 10.21): The recorded talk block on the right, the slide set block on the left and the information block on the bottom. The recorded talk is synchronized with the slides extracted from a presentation slide set. The navigation banner at the bottom of the slide set block allows the user to select any slide from the presentation in order to get the corresponding talk video sequence played. Using SMAC to replay the recorded talks provide users a simple and easy navigation

Figure 10.21: SMAC - visualization module - talk video recording replay interface - I

through the different sequences of the talk. In addition, the users may display the aggregated annotations provided by the life cycle of the conference. The users have the possibility to search within the talks of the same session of the current talk (Fig. 10.22, part -1-). The interface offers also a keyword search within the current talk (Fig. 10.22, part -2-), the engine searches within the metadata extracted from the file presentation, or from the speech-to-text module or from the CALISEMA annotation or any other information available. At any time the users can go back to the search interface (INVENIO, NAVIR) to perform another search.

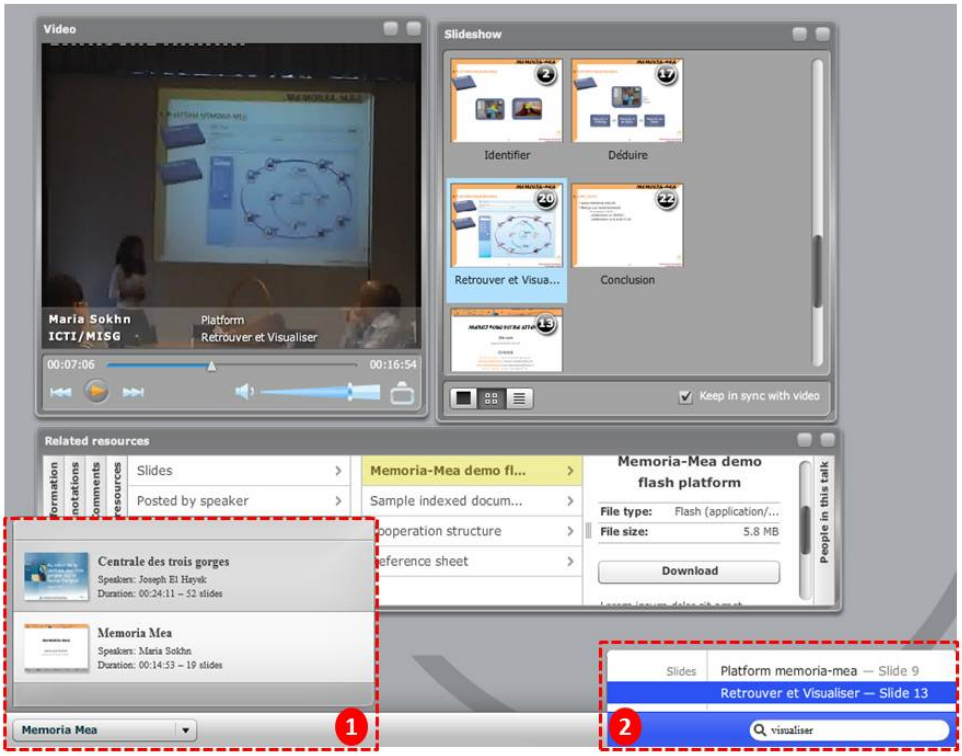


Figure 10.22: SMAC - visualization module - talk video recording replay interface - II

10.3 Platform Network

Scientific conferences being a gathering of individuals and members, this contributes to the existing of distributed information and the need of sharing and accessing resources. The large scale of the scientific community can be exploited by using the P2P potential. As discussed in the state of the art, few frameworks, in the context of the content based multimedia retrieval, made use of the richness carried by the peers in a P2P architecture. We intend with the framework CALIMERA to make use of the P2P network. We choose a structured decentralized network architecture where some of the peers may be considered as super peers referred in this thesis work as the *calimera-peers* e.g. CALIMERA EIAFR, CALIMERA ESIB that integrates the conference life cycle set of tools, or the *normal-peers* that represent the users that might want to share, annotation or search for resources (Fig. 10.23). Two proof of concept were developed to be tested, one using CHORD and one developed over the JXTA framework. In this thesis we present the work done over the JXTA framework. This choice was made thanks to the easy handling of the framework and to the possibility it offers regarding the medium that can be connected (PDA, computer, phone, etc.). On the top of the framework JXTA we integrated the RDF technology in order to allow a content based description of the available multimedia resources. The JXTA framework is a series of classes and methods for managing and transmitting application and controlling data between JXTA compatible peers platform.

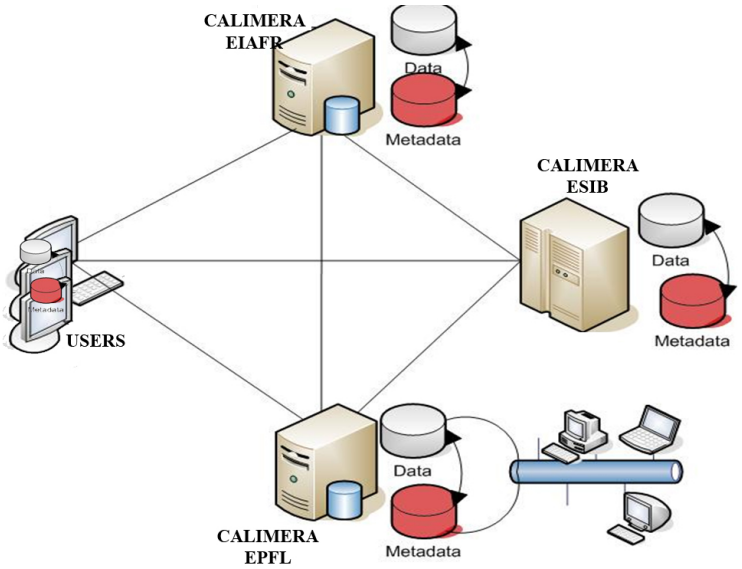


Figure 10.23: Platform Network

When using the CALIMERA platform, the users can perform the following tasks: Share multimedia resources, submit related data, (manually and semi-

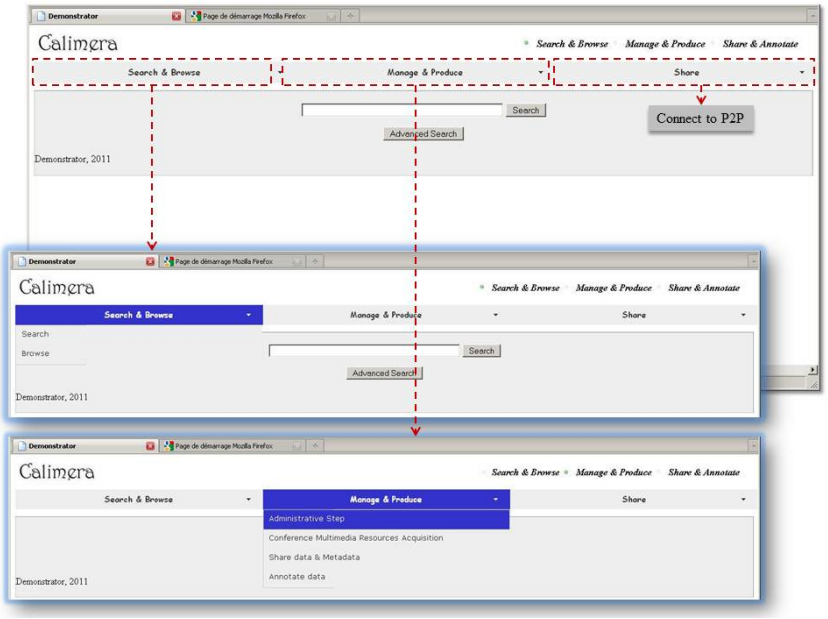


Figure 10.24: CALIMERA platform - main page

manually annotate the data based on the designed conference model HELO, and initiate faceted and traditional keyword search.

When the P2P application is launched, the peer can connect to the JXTA network peer group then send advertisements to the peers on the network to exchange some information about their respective sockets. The class JXTAManager is responsible for connecting the peer to the network. It uses the NetworkManager class provided by JXTA framework and creates a multicast JXTA socket on each peer, allowing it to send and receive requests and responses. The application makes use of the Discovery Service offered by JXTA in order to periodically detect the new peers in the network and connect to them. The proposed application allows any peer to join the network (*normal-peers* and *calimera-peers*).

The users can have access to the P2P platform through an application composed of different modules: (1) Graphical User Interface (GUI): It allows the user to interact with the application by sharing, submitting and managing the data resources and performing search tasks. (2) Controller: It collaborates with the user interface to send the performed search to the search manager and manage the ontology using Jena, a Java framework for building semantic web applications JENA. (3) Search Manager: A module between the application modules (GUI, Controller, Jena) and the network modules (JXTA manager, Download manager - described hereafter). (4) JXTA Manager: It is responsible of the connections and the discovery of new peers in the network. (5) Jena: The programming toolkit that manages the tasks related to the semantic annotation and search. (6)

Download Manager: It is responsible of the upload and download of the different data resources. When the users are connected to the P2P platform (Fig. 10.24), the search is performed over the peers neighbors which in turn search in the neighboring peers (Fig. 10.25, 10.26).

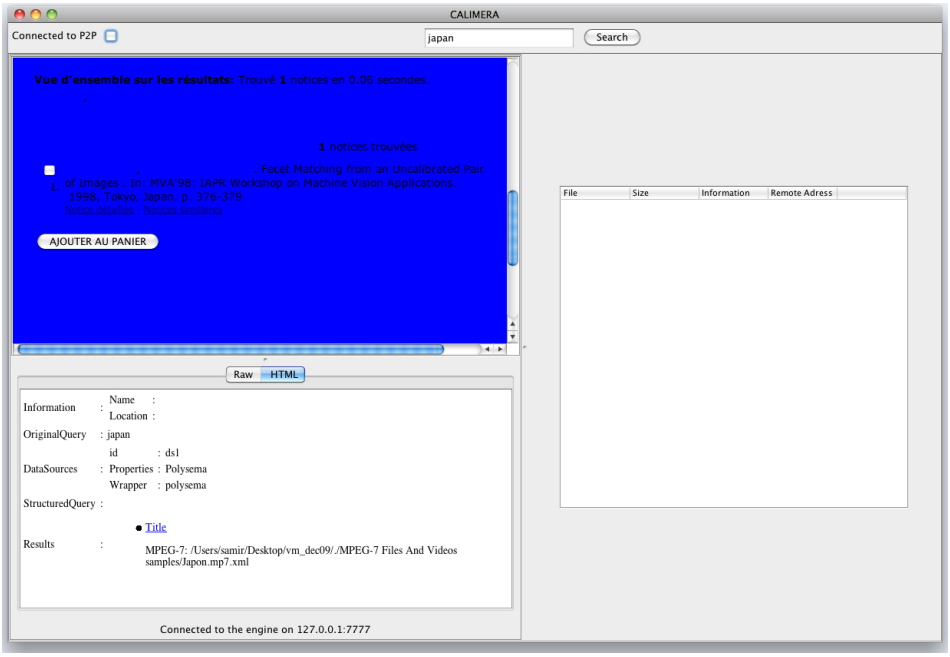


Figure 10.25: CALIMERA platform - example of a user request without connected peers

10.4 Platform Testbed and Evaluation

In order to validate the proposed approach we setup a testbed: Initially, the first tests mainly for video analysis and features extraction, were performed over a set of videos conference provided by the CERN. Afterwards, we build this thesis testbed over different types of events namely internship presentations, diploma work presentations, master thesis presentations, academic courses and the most important was the international scientific conference (Atlantic Web Intelligence Conference 2011 that took place in Fribourg - AWIC2011 [AWI]). The overall testbed contains about fifty hours of recorded talks, it contains the corresponding slides and other data resources such as pictures and papers which are stored in the system and the peers. The testbed contains also the recordings annotation provided either manually (through the annotation tools, RDF and OWL being the main output format) or automatically (through the features extraction modules, MPEG7 being the main output format) (Fig. 10.27).

The state of the art presented in PART I, showed the lacks and open issues

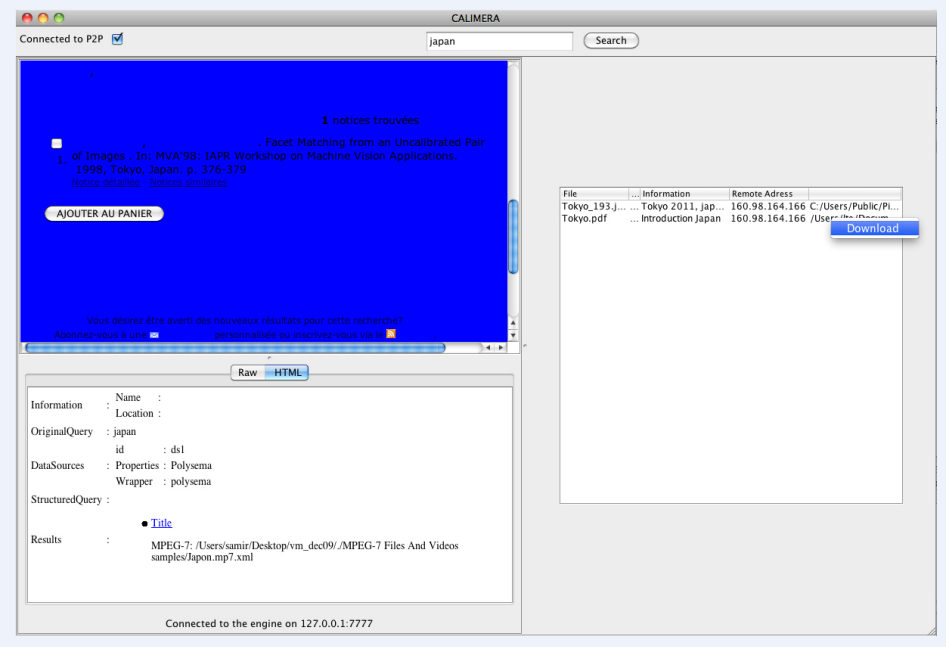


Figure 10.26: CALIMERA platform - example of a user request with connected peers

Figure 10.27: CALIMERA testbed

in the domain of multimedia information retrieval. By implementing CALIMERA a number of the raised issues were addressed achieving improvements in this domain. As mentioned earlier, the success of a multimedia retrieval system is mainly determined by its ability to support users requirements in a specific domain. To that end we decided to design the proposed framework based on a the scientific and academic conference domain. In order to support an effective management and reuse of the knowledge conveyed within the conference life cycle we proposed HELO, an ontological model. This model integrates the information related to the conference life cycle, hence, enriching the description related to a video talk recording. The evaluation of the different approaches integrated in the framework namely the ontological model, the graphical user interfaces for both annotation and retrieval, are of great interest and importance in order to improve the ontology driven framework proposed in this thesis. This evaluation should exploit the following axes:

- **The evaluation of the ontological model HELO**: This evaluation should validate the coherence of the chosen scopes and their consistency according to different point of view. It should also measure the ease of learning and use of this model. the evaluation could be done through a set of question addressed to the users.

- **The evaluation of the semantic annotation interface**: This evaluation should verify the effectiveness of the user interfaces and the added value of the integration of conference video segmentation. This evaluation could be done through a comparison between traditional annotation system and the proposed one which have the advantage of being integrated within a conference framework. Criteria such as the simplicity of the interface, the time consuming, etc. should be taken into account.

- **The evaluation of the semantic search engine interface**: This evaluation should consider and verify, through a set of experiments and concrete examples, the expressiveness and effectiveness of the graphical user interfaces used for querying and browsing. This evaluation could be done through a comparison between the seek of the same information through a traditional system and the one we propose. Criteria such as the simplicity of the query interface, the time needed to get the result, the pertinence of the results, etc. should be taken into account.

- **The evaluation of the multimedia rendering and browsing interfaces**: This evaluation should evaluate the need, the pertinence and the advantages of the proposed navigation and browsing interfaces. The evaluation should compare the proposed interfaces to existing ones at different level of retrieval.

- **The evaluation of the P2P Network**: This evaluation should consider the advantage gained by integrated the P2P network for both data and metadata management mainly the model based annotation aspect and the performance

of some modules if distributed over the peers such as the module of speech-to-text.

10.5 Conclusion

Severals prototypes were integrated into the framework in order to validate the proposed approach. The integrated tools cover the life cycle of a conference. The automation of the conference life cycle process through these tools offers the users advantages in terms of time by promoting the use of existing information and the integration of automatically extracted information without a human intervention. In addition to the data and metadata management, we focused in this thesis prototype on the ontology driven browsing and visualization approaches.

Part V

CONCLUSION AND PERSPECTIVES

Chapter 11

Conclusion and Perspectives

Conclusion

The recent technological advances have made a large amount of multimedia resources available to users in a variety of areas. However, without adequate approaches for effective access to multimedia content, this large and valuable body of data is barely accessible and in effect unusable. In this thesis we aimed at providing a novel and valuable approach for content based multimedia retrieval for scientific and academic events. The literature review showed that the scientific and academic events are gradually increasing that the quantity of high quality scientific information disseminated via conferences is growing steadily and that a large number of these events are being recorded for subsequent use (replay, research, course preparation, and so on). Many characteristics specific to the scientific events can be identified and handled in order to enhance the content based multimedia retrieval.

The overall analysis of the state of the art led us to conclude with a set of observations and reflections which motivated the main research axes of the work proposed in this thesis. In this thesis we aimed at providing a novel approach for content based multimedia retrieval. We explored semantic approaches to content based management browsing and visualization of the multimedia resources generated for and during scientific conferences. We aimed at providing users an integrated framework that enhances the conference multimedia information retrieval by bridging the so-called "semantic gap". The gap which exists between the explicit knowledge representation required by users who search the multimedia resources and the implicit knowledge conveyed within a conference life cycle. Three main issues were addressed and the main contribution were focused towards these issues:

- **The first issue is the knowledge modeling**: We designed an ontology model that was aimed at being used as an underlying semantic model for the conference multimedia retrieval process. The state of the art study showed that an increasing attention is given to ontology based approach, as it offers promising solutions to enhance and allow information retrieval at a semantic level. Existing models cover some of the required structures and content based descriptions in the domain of scientific conferences. We designed and implemented HELO, an ontology that models the information conveyed during the life cycle of a scientific conference. The model is composed of granular concepts that are representative of users' annotation, and of eight different

concepts which we call *scopes* that are representative of users' browsing and visualization processes.

- **The second addressed issue is the knowledge management**: In this thesis we aimed at improving the content based multimedia description by aggregating available information generated through the life cycle of a conference. To address this issue, we designed an integrated approach for content based information retrieval for the domain of scientific events. The approach is based the ontology of conferences. We aimed at making use of the existing unused information conveyed during a conference life cycle by integrating this information in to HELO, for subsequent use, throughout the conference life process.

- **The third addressed issue is the knowledge browsing and visualization**: We aimed at improving the multimedia retrieval by offering a model based browsing approach. We proposed a browsing and visualization approach based on HELO that takes into account the life cycle of information retrieval. We aimed at giving the users the ability to explore the conference information with respect to the *scopes* defined based on their requirements.

In order to unify the many efforts that have been conducted during the thesis by integrating novel approaches within the multimedia information retrieval process (namely annotation, browsing and visualization) we designed an integrated end-to-end approach for conference multimedia retrieval. This approach, makes use of the life cycle of a multimedia resources thanks to the underlying semantic model HELO. We aimed in this thesis approach at making use of the potential of P2P networks mainly for content based annotation and retrieval. To validate this thesis proposed approach we set up CALIMERA, an ontology driven framework which integrates a set of tools for the retrieval of conference multimedia resources. A set of domain specific approaches (tools, algorithms, and interfaces) were developed.

At the annotation level, the thesis work presents a significant improvement in the conference video recording annotation: (1) We made use of existing information which was not used in previous approaches. (2) We designed and implemented conference video segmentation algorithm. The proposed algorithm relies on slides change detection. This was based on the fact that users may need to visualize a specific video sequence of a conference. This sequence in most cases corresponds to a specific slide or set of slides. (3) We proposed an automatic annotation through a speech-to-text algorithm which automatically assigns time based annotation. (4) We offered users an enhanced tool, designed for conference video annotation. This tool is a semi-manual annotation tools based on HELO a common ontology model.

At the browsing and visualization level, the thesis work presents a novel approach of searching and visualizing conferences based on a conference ontology. (1) We offer the users the possibility to navigate through the multimedia resources

based on the semantic model. Using this browsing approach, referred in this thesis work as the faceted search, allowed the users to explore the available information and to visualize the search result according to different scopes such as *PersonScope*, *CommunityScope*, *TemporalScope*, and so on. (2) We also offered the users the possibility to filter the results according to more than one Scope. This approach is essential and interesting due to the data filtering aspects which, depending on the cases, may further enhance the visualization of the search result.

Finally, CALIMERA presents an end-to-end integrated framework, which may be used as a useful basis for experiments, mainly for an operational platform but more probably as a first step for users evaluation. The current results obtained in this thesis work call for various future research directions. In the next section we propose a few of such improvement and extension possibilities.

Perspectives

One of the imminent task to be done is the evaluation of the proposed platform particularly when performed over the P2P network over a full-scale data set. As mentioned earlier, the success of a multimedia retrieval system is mainly determined by its ability to support users requirements. To that end we decided to design the proposed framework based on a specific domain, aiming by this at focusing on specific user requirements. The evaluation of the different features integrated in the framework namely the model, the graphical user interfaces both for annotation and retrieval are of great interest to improve the proposed approach and consequently the CALIMERA platform. This evaluation should exploit the evaluation of the ontological model HELO, the evaluation of the semantic annotation interface, the evaluation of the semantic search engine interface, and the evaluation of the multimedia rendering and browsing interfaces.

As next steps, it would be of interest to enhance the proposed approach. This could be done at different levels: (1) In the area of the data and metadata management, enhancements may be conveyed over the content extraction field. Indeed, many mid-level features, such as face recognition, can be extracted which can potentially be used for indexing and retrieval. Semantic annotation techniques such as video Optical Character Recognition (OCR), objects tracking can be also integrated into the platform to enrich the available annotation. (2) In the area of retrieval, the virtual query approach should be developed and evaluate. Further user interfaces should support data insertion, update and relevance feedback. The proposed user interfaces should be completed by implementing the ones that have not been developed yet. (3) In the P2P network area, an important and interesting aspect may be brought to improve the performance of the platform. As the proposed platform is modular we can distribute the different modules over the connected peers making use of their Control Process Unit (CPU). This distribution

is interesting for the speech-to-text module and the recording tool primarily when we have parallel sessions that should be recorded.

Other important aspects could be taken into account in this area of work and may be considered as an extension of the proposed platform. This may include the extension of the model HELO including other existing ontologies and making use of the large amount of information that we may find on the web. Information such as the scientific literature digital libraries and information that we can extract from current social networks and P2P network.

Part VI

Bibliography

Bibliography

[Abello 2006] James Abello, Frank Van Ham and Neeraj Krishnan. *ASK-GraphView: A Large Scale Graph Visualization System.* IEEE Transactions on Visualization and Computer Graphics, page 2006, 2006. (Cited on page 33.)

[Aberer 2005] K. Aberer, L.O. Alima, A. Ghodsi, S. Girdzijauskas, S. Haridi and M. Hauswirth. *The essence of P2P: a reference architecture for overlay networks.* In Peer-to-Peer Computing, 2005. P2P 2005. Fifth IEEE International Conference on, pages 11 – 20, 2005. (Cited on page 51.)

[Aberer 2011] Karl Aberer. *Peer-to-Peer Data Management.* 2011. (Cited on page 38.)

[ACM] *ACM Taxonomy* - *http://www.computer.org/portal/web/publications/acmtaxonomy*. (Cited on pages 65 and 118.)

[ADV] *ADVENE - Annotate Digital Video, Exchange on the Net* - *http://liris.cnrs.fr/advene/*. (Cited on page 26.)

[Akrivas 2007] Giorgos Akrivas, Georgios Th. Papadopoulos, Matthijs Douze, Johannes Heinecke, Noel Connor, Carsten Saathoff and Simon Waddington. *Knowledge-based Semantic Annotation and Retrieval of Multimedia Content.* In 2nd International Conference on Semantic and Digital Media Technologies, 2007. (Cited on page 25.)

[AKT] *Advanced Knowledge Technology (AKT) ontology* - *http://www.aktors.org/publications/ontology/*. (Cited on page 21.)

[AMA] *Amazone online shopping* - *http://www.amazon.com/*. (Cited on page 37.)

[ANN] *Annomation - video annotation tool* - *http://annomation.open.ac.uk/annomation*. (Cited on page 26.)

[Antoniou 2004] G. Antoniou and F. van Harmelen. A semantic web primer. The MIT Press, April 2004. (Cited on page 18.)

[ANV] *Anvil - video annotation tool* - *http://www.anvil-software.de/*. (Cited on page 26.)

[ARN] *Arnetminer - expertise oriented search system* - *http://www.arnetminer.org/*. (Cited on page 34.)

[Atitallah 2011a] Samir Atitallah, Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Knowledge management in next generation networks - Semantic multimedia management in P2P Network.* 2011. (Cited on page 6.)

[Atitallah 2011b] Samir Atitallah, Tobias Wunden, Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Speech Indexation in REPLAY.* In AWIC, pages 111–119, 2011. (Cited on page 6.)

[AWI] *Atlantic Web Intelligence Conference (AWIC) - http://awic2011.eia-fr.ch/.* (Cited on page 128.)

[B. Erol 2003] J. Hull B. Erol and D.-S. Lee. *Linking multimedia presentations with their symbolic source documents: algorithm and applications.* pages 498–507. eleventh ACM international conference on Multimedia, 2003. (Cited on page 108.)

[Baeza-Yates 1999] Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. Modern information retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999. (Cited on page 32.)

[Bagdanov 2007] Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo, Giuseppe Serra and Carlo Torniai. *Semantic annotation and retrieval of video events using multimedia ontologies.* In Proceedings of the International Conference on Semantic Computing, pages 713–720, Washington, DC, USA, 2007. IEEE Computer Society. (Cited on page 33.)

[Bao 2004] Jie Bao, Yu Cao, Wallapak Tavanapong, , Vasant Honavar and Vasant Honavar. *Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation.* In In: International Conference on Information and Knowledge Engineering (IKE 04). CSREA, pages 82–88. Press, 2004. (Cited on page 25.)

[Barre 2006] Kathryn La Barre. *PhD Thesis: "The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design".* PhD thesis, Indiana University, 2006. (Cited on page 29.)

[Behera 2004] Ardhendu Behera, Denis Lalanne and Rolf Ingold. *Looking at projected documents: event detection and document identification.* In IEEE international conference on multimedia and expo, pages 2127–2130, vol.3, June 2004. (Cited on page 72.)

[Berners-Lee 2001] Tim Berners-Lee, James Hendler and Ora Lassila. *The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.* 2001. (Cited on page 18.)

[BIB] *A reference management software for formatting lists of references - http://www.bibtex.org.* (Cited on page 22.)

[Blanco 2006] R. Blanco, N. Ahmed, D. Hadaller, L. G. A. Sung, H. Li and M. A. Soliman. *A Survey of Data Management in Peer-to-Peer Systems.* Rapport technique, 2006. (Cited on page 48.)

[Boreczky 1998] John S. Boreczky and Lynn D. Wilcox. *A hidden Markov model framework for video segmentation using audioand image features.* In IEEE international conference on acoustics, speech and signal processing, 1998., pages 3741–3744, vol.6, 1998. (Cited on page 71.)

[Boszormenyi L. 2005] Kropfberger M. Boszormenyi L. Muller A. *Annotation and presentation of content-variations in a webbased search environment for video.* pages 67–74. Istvan Simonics (Hrsg.): Technology-enhanced Learning with Ubiquitous Applications of Integrated Web, Digital TV and Mobile Technologie, July 2005. (Cited on page 25.)

[Burkhard 2005] Remo Burkhard. *Towards a Framework and a Model for Knowledge Visualization: Synergies Between Information and Knowledge Visualization.* In Sigmar-Olaf Tergan and Tanja Keller, editeurs, Knowledge and Information Visualization, volume 3426 of *Lecture Notes in Computer Science*, pages 29–42. Springer Berlin / Heidelberg, 2005. (Cited on page 31.)

[Cai 2004] Min Cai and Martin Frank. *RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network.* In Proceedings of the 13th international conference on World Wide Web, WWW '04, pages 650–657. ACM, 2004. (Cited on page 48.)

[Carbonaro 2008] Antonella Carbonaro. *An Integrated and Interactive Video Retrieval Framework with Hierarchical Learning Models and Semantic Clustering Strategy.* 2008. (Cited on page 44.)

[Chang 1998] Shih-Fu Chang, W. Chen and H. Sundaram. *VideoQ: a fully automated video retrieval system using motion sketches.* In Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop on, pages 270 –271, 1998. (Cited on page 37.)

[Chawathe 1994] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman and J. Widom. *The TSIMMIS Project: Integration of Heterogenous Information Sources.* In Information Processing Society of Japan (IPSJ 1994), 1994. (Cited on page 39.)

[Chen 2009] Min Chen, David Ebert, Hans Hagen, Robert S. Laramee, Robert van Liere, Kwan-Liu Ma, William Ribarsky, Gerik Scheuermann and Deborah Silver. *Data, Information, and Knowledge in Visualization.* IEEE Comput. Graph. Appl., vol. 29, pages 12–19, 2009. (Cited on page 31.)

[Chiu 2000] Patrick Chiu, Jonathan Foote, Andreas Girgensohn and John Boreczky. *Automatically linking multimedia meeting documents by image Matching.* In Proceedings of the eleventh ACM on Hypertext and hypermedia, pages 244–245, 2000. (Cited on page 72.)

[COF a] *BusinessDictionary - conference definition.* (Cited on page 4.)

[COF b] *Event Research and Investigation Services - conference definition* - http://www.event-research.com/event.php. (Cited on page 4.)

[COF c] *Oxford Dictionaries - conference definition* - http://oxforddictionaries.com/definition/conference?region=us. (Cited on page 4.)

[Correia 1999] Nuno Correia and Teresa Chambel. *Active video watching using annotation.* In Proceedings of the seventh ACM international conference on Multimedia (Part 2), MULTIMEDIA '99, pages 151–154. ACM, 1999. (Cited on page 25.)

[CRE] *Crescendo - digital dictation recognition* - http://www.crescendosystems.co.uk/. (Cited on page 76.)

[Cudre-Mauroux 2007] P. Cudre-Mauroux, S. Agarwal and K. Aberer. *GridVine: An Infrastructure for Peer Information Management.* Internet Computing, IEEE, vol. 11, no. 5, pages 36 –44, sept.-oct. 2007. (Cited on page 48.)

[Cullot 2003] Nadine Cullot, Christine Parent, Stefano Spaccapietra and Christelle Vangenot. *Ontologies: A Contribution to the DL/DB debate.* 2003. (Cited on page 19.)

[Dakka 2005] Wisam Dakka, Panagiotis G. Ipeirotis and Kenneth R. Wood. *Automatic construction of multifaceted browsing interfaces.* In Proceedings of the 14th ACM international conference on Information and knowledge management, CIKM '05, pages 768–775. ACM, 2005. (Cited on page 31.)

[DBL] *Digital Bibliography and Library Project (DBLP)* - http://www.informatik.uni-trier.de/~ley/db/. (Cited on page 34.)

[Dimitrova 2002] Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan, Thomas Huang and Avideh Zakhor. *Applications of Video-Content Analysis and Retrieval.* IEEE MultiMedia, vol. 9, pages 42–55, 2002. (Cited on page 24.)

[DIR] *The MS-DirectShow API* - http://msdn.microsoft.com/en-us/library/ms783323(VS.85).aspx. (Cited on page 105.)

[Domenig 2000] Ruxandra Domenig and Klaus R. Dittrich. *A query based approach for integrating heterogeneous data sources.* In Proceedings of the ninth international conference on Information and knowledge management, CIKM '00, pages 453–460. ACM, 2000. (Cited on page 39.)

[Dong 2006] Aijuan Dong and Honglin Li. *Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval.* In Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing - Vol 2 - Workshops - Volume 02, pages 158–165. IEEE Computer Society, 2006. (Cited on pages 24 and 25.)

[DRA a] *DRAGO* http://drago.itc.it/index.html. (Cited on page 48.)

[DRA b] *Dragon - speech recognition software -* http://www.nuance.com/dragon/index.htm. (Cited on page 76.)

[Dumoulin 2010] Joel Dumoulin, Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Poster: Multimedia information browsing and visualization.* 2010. (Cited on page 6.)

[Dumoulin 2011] Joel Dumoulin, Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Vers une approche de visualization multivues dans un systeme d'information repartie.* 2011. (Cited on page 6.)

[EBA] *EBay online shopping -* http://www.ebay.com/. (Cited on page 37.)

[EBI] *eBiquity Conference Ontology -* http://ebiquity.umbc.edu/ontology/. (Cited on page 21.)

[Edgington 2004] Theresa Edgington, Beomjin Choi, Katherine Henson, T.S. Raghu and Ajay Vinze. *Adopting ontology to facilitate knowledge sharing.* Commun. ACM, vol. 47, pages 85–90, November 2004. (Cited on page 19.)

[EDU] (Cited on page 51.)

[ELA] *Educational LANguage (ELAN) -* http://www.mpi.nl/tools/elan.html. (Cited on page 26.)

[Erol 2003] Berna Erol, Jonathan J. Hull and Dar-Shyang Lee. *Linking multimedia presentations with their symbolic source documents: algorithm and applications.* In Proceedings of the eleventh ACM international conference on Multimedia, pages 498–507, 2003. (Cited on page 72.)

[ESW] *European Semantic Web Conference(ESWC) ontology -* http://www.eswc2006.org/technologies/ontology-content/2006-09-21.rdf. (Cited on page 21.)

[EXM] *EXtensible MARkup Language for Discourse Annotation (EXMARaLDA) -* http://www.rrz.unihamburg.de/exmaralda/index-en.html. (Cited on page 26.)

[FAG] *Usability Studies of Faceted Browsing: A Literature Review.* vol. 29. (Cited on page 29.)

[Fan 2006] Quanfu Fan, Kobus Barnard, Arnon Amir, Alon Efrat and Ming Lin. *Matching Slides To Presentation Videos Using SIFT Scene Background Matching.* In POSTER SESSION: Poster session 2: annotation, summarization, and visualization, pages 239–248, 2006. (Cited on page 72.)

[Fatemi 2003] Nastaran Fatemi. *PhD Thesis: "A semantic views model for audio-visual indexing and retrieval"*. PhD thesis, Ecole Polytechnique Federale de Lausanne, 2003. (Cited on pages 31, 38 and 44.)

[FLA a] *Flamenco Search Interface project.* (Cited on page 29.)

[FLA b] *Flare library -* `http://flare.prefuse.org`. (Cited on page 118.)

[FOA] *Friend Of A Friend ontology -* `http://www.foaf-project.org/`. (Cited on pages 22 and 64.)

[GAR] *Gartner Identifies Top Ten Disruptive Technologies for 2008 to 2012 -* `http://www.gartner.com/it/page.jsp?id=681107`. (Cited on page 18.)

[GLA] *Was Ranganathan a Yahoo!? InterNIC News In End User's Corner column -* `http://www.scout.cs.wisc.edu/addserv/toolkit/enduser/archive/1998/euc-9803.html`. (Cited on page 31.)

[Glaser 2007] Hugh Glaser and Ian C. Millard. *RKB Explorer: Application and Infrastructure.* In International Semantic Web Conference, 2007. (Cited on pages 34 and 44.)

[Glaser 2008] Hugh Glaser, Ian C. Millard and Afraz Jaffri. *RKBExplorer.com: a knowledge driven infrastructure for linked data providers.* In Proceedings of the 5th European semantic web conference on The semantic web: research and applications, ESWC'08, pages 797–801. Springer-Verlag, 2008. (Cited on page 44.)

[Gong 2001] Li Gong. *JXTA: A Network Programming Environment.* IEEE Internet Computing, vol. 5, pages 88–95, 2001. (Cited on page 49.)

[GRA] *RDF Gravity - RDF Graph Visualization Tool -* `http://semweb.salzburgresearch.at/apps/rdf-gravity/`. (Cited on page 33.)

[H. 2006] Weihong H. and Alain M. *ConKMeL: A contextual knowledge management framework to support multimedia e-Learning.* In Multimedia Tools and Applications, volume 30, pages 205–219, August 2006. (Cited on page 44.)

[Handschuh 2003] Siegfried Handschuh and Steffen Staab. *CREAM: CREAting Metadata for the Semantic Web.* Computer Networks, vol. 42, no. 5, pages 579–598, 2003. (Cited on page 26.)

[Hao Ding] Ingeborg Torvik Solvberg Hao Ding. *Semantic Search in Peer-to-Peer based Digital Libraries.* (Cited on page 43.)

[HAY] *Haystack Project - http://groups.csail.mit.edu/haystack/.* (Cited on page 41.)

[Hearst 2006] Marti A. Hearst. *Clustering versus faceted categories for information exploration.* Commun. ACM, vol. 49, pages 59–61, 2006. (Cited on page 29.)

[Hellmann 2010] Sebastian Hellmann. *The Semantic Gap of Formalized Meaning.* In ESWC (2)'10, pages 462–466, 2010. (Cited on page 3.)

[Hirsch 2009] C. Hirsch, J. Hosking, J. Grundy, T. Chaffe, D. MacDonald and Y. Halytskyy. *The Visual Wiki: A New Metaphor for Knowledge Access and Management.* In System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on, pages 1 –10, 2009. (Cited on page 32.)

[Hoang 2006] Hanh Huu Hoang, Amin Andjomshoaa and A Min Tjoa. *Towards a New Approach for Information Retrieval in the SemanticLIFE Digital Memory Framework.* Web Intelligence, IEEE / WIC / ACM International Conference on, vol. 0, pages 485–488, 2006. (Cited on page 39.)

[Huynh 2007] David F. Huynh, David R. Karger and Robert C. Miller. *Exhibit: lightweight structured data publishing.* In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 737–746. ACM, 2007. (Cited on page 31.)

[Hyvonen 2002] Eero Hyvonen and Avril Styrman. *Ontology-based image retrieval.* pages 15–27, 2002. (Cited on page 25.)

[IFI] *iFinder - an MPEG7-based retrieval system for distributed multimedia content tool - `http://www.imk.fhg.de/de/ifinder`.* (Cited on page 26.)

[INF] *javascript InfoVis Toolkit - `http://thejit.org`.* (Cited on page 118.)

[INV] *Invenio digital library - modules overview - `http://invenio-demo.cern.ch/help/hacking/modules-overview`.* (Cited on page 110.)

[ISA] *IsaViz - a visual authoring tool for RDF - .* (Cited on page 33.)

[Izza 2006] Said Izza, Lucien Vincent and Patrick Burlat. *A Unified Framework for Enterprise Integration An Ontology-Driven Service-Oriented Approach*, 2006. (Cited on page 38.)

[JOA] *JOANNEUM research group - `http://mpeg-7.joanneum.at`.* (Cited on page 26.)

[John K. 1997] Rick K. John K. *Accessing Multimedia through Concept Clustering.* pages 19–26. ACM SIGCHI, 1997. (Cited on page 108.)

[JXT] *JXTA - a language and platform independent protocol for P2P networking - `http://www.jxta.org`.* (Cited on page 49.)

[Keim 2002] Daniel A. Keim. *Information Visualization and Visual Data Mining.* IEEE Transactions on Visualization and Computer Graphics, vol. 8, pages 1–8, 2002. (Cited on page 33.)

[Kherfi 2004] M. L. Kherfi, D. Ziou and A. Bernardi. *Image Retrieval from the World Wide Web: Issues, Techniques, and Systems.* ACM Comput. Surv., vol. 36, pages 35–67, 2004. (Cited on page 28.)

[Kilany 2008] R. Kilany, M. Chamoun and A. Serhrouchni. *GSS: A generic P2P semantic service oriented architecture.* In Software, Telecommunications and Computer Networks, 2008. SoftCOM 2008. 16th International Conference on, pages 409 –414, 2008. (Cited on page 48.)

[Kiryakov 2001] Atanas Kiryakov, Kiril Iv. Simov and Marin Dimitrov. *OntoMap: portal for upper-level ontologies.* FOIS '01, pages 47–58. ACM, 2001. (Cited on page 19.)

[Kitamura 2003] Yoshinobu Kitamura and Riichiro Mizoguchi. *An Ontological Schema for Sharing Conceptual Engineering.* In Knowledge," International Workshop on Semantic Web Foundations and Application Technologies, pages 25–28, 2003. (Cited on page 18.)

[Klampanos 2003] Iraklis A. Klampanos and Joemon M. Jose. *An architecture for peer-to-peer information retrieval.* In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, SIGIR '03, pages 401–402. ACM, 2003. (Cited on pages 43 and 48.)

[Klette 1991] Reinhard Klette and Ivan Stojmenovic'. *M.Tamer Ozsu and Patrick Valduriez, Principles of distributed database systems.* In Graphical Models Image Processing, pages 295–301. Prentice-Hall, 1991. (Cited on page 38.)

[Knud Moller 2007] Siegfried Handschuh John Domingue Knud Moller Tom Heath. *Recipes for Semantic Web Dog Food - The ESWC and ISWC Metadata Projects.* In ISWC/ASWC, pages 802–815, 2007. (Cited on pages 21, 22 and 23.)

[Koren 2007] Jonathan Koren, Andrew Leung, Yi Zhang, Carlos Maltzahn, Sasha Ames and Ethan Miller. *Searching and navigating petabyte-scale file systems based on facets.* In Proceedings of the 2nd international workshop on Petascale data storage: held in conjunction with Supercomputing '07, PDSW '07, pages 21–25. ACM, 2007. (Cited on page 31.)

[La Cascia 1996] M. La Cascia and E. Ardizzone. *JACOB: just a content-based query system for video databases.* In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, volume 2, pages 1216 –1219 vol. 2, 1996. (Cited on page 37.)

[Lambert 2010] Dave Lambert and Hong Qing Yu. *Linked Data Based Video Annotation and Browsing for Distance Learning.* In The 2nd International Workshop on Semantic Web Applications in Higher Education (SemHE'10), November 2010. (Cited on page 26.)

[Lawson 2004] K. Lawson. *Visualizing the semantic web: XML-based internet and information visualization [Book Review]*. Distributed Systems Online, IEEE, vol. 5, no. 2, pages 6/1 – 6/3, feb. 2004. (Cited on page 25.)

[Lee 2009] Bongshin Lee, Greg Smith, George G. Robertson, Mary Czerwinski and Desney S. Tan. *FacetLens: exposing trends and relationships to support sensemaking within faceted datasets*. In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, pages 1293–1302, New York, NY, USA, 2009. ACM. (Cited on page 31.)

[Li 1995] John Z. Li, M. Tamer Ozsu and Duane Szafron. *Query Languages in Multimedia Database Systems*. Rapport technique, 1995. (Cited on page 28.)

[Lienhart 1996] Rainer Lienhart. *Comparison of automatic shot boundary detection algorithms*. In Storage and Retrieval for Still Image and Video Databases IV, pages 170–179, 1996. (Cited on page 71.)

[LIN] *Linked Data - http://www.linkeddata.org/*. (Cited on page 18.)

[Liu 2007] Ying Liu, Dengsheng Zhang, Guojun Lu and Wei-Ying Ma. *A survey of content-based image retrieval with high-level semantics*. Pattern Recogn., vol. 40, pages 262–282, 2007. (Cited on page 3.)

[Lorenzo Corsani] Pilu Crescenzi Leonardo Lanzi Lorenzo Corsani Carlo Nocentini. (Cited on page 51.)

[Lua 2005] Eng Keong Lua, Jon Crowcroft, Marcelo Pias, Ravi Sharma and Steven Lim. *A Survey and Comparison of Peer-to-Peer Overlay Network Schemes*. IEEE Communications Surveys and Tutorials, vol. 7, pages 72–93, 2005. (Cited on page 48.)

[Luo 2008a] Hangzai Luo, Jiahang Yang, Aoying Zhou, Jianping Fan and Tianming Hu. *Knowledge Mining and Visualization on News Webpages and Large-Scale News Video Database*. In Proceedings of the 2008 IFIP International Conference on Network and Parallel Computing, pages 452–459, Washington, DC, USA, 2008. IEEE Computer Society. (Cited on page 33.)

[Luo 2008b] Hangzai Luo, Jiahang Yang, Aoying Zhou, Jianping Fan and Tianming Hu. *Knowledge Mining and Visualization on News Webpages and Large-Scale News Video Database*. In Network and Parallel Computing, 2008. NPC 2008. IFIP International Conference on, pages 452 –459, 2008. (Cited on page 33.)

[Lux 2003] Mathias Lux and al. *Caliph & emir: Semantic annotation and retrieval*. In Personal Digital Photo Libraries, pages 85–89. CAiSE '03 Forum at 15th Conference on Advanced Information Systems Engineering, 2003. (Cited on page 25.)

[Ma 2010] Hao Ma, Jianke Zhu, Michael Rung-Tsong Lyu and Irwin King. *Bridging the Semantic Gap Between Image Contents and Tags*. IEEE Transactions on Multimedia, vol. 12, pages 462–473, 2010. (Cited on page 3.)

[Marchand-Maillet 2000] Stephane Marchand-Maillet. *Content-based Video Retrieval: An overview*. 2000. (Cited on pages 37 and 38.)

[MAT] *Matterhorn Project, open-source platform to support the management of educational audio and video content*. Opencast community. http://www.opencastproject.org/project/matterhorn. (Cited on page 44.)

[Mauroux 2006] Philippe Cudre Mauroux. *Emergent semantics: rethinking interoperability for large scale decentralized information systems*. 2006. (Cited on page 38.)

[Maynard 2003] Diana Maynard. *Multi-Source and Multilingual Information Extraction*. 2003. (Cited on page 24.)

[m.c. schraefel 2005] m.c. schraefel, Daniel alexander Smith, Alisdair Owens, Alistair Russell, Craig Harris and Max L. Wilson. *The evolving mSpace platform: leveraging the Semantic Web on the Trail of the Memex*. In Hypertext, 2005, pages 174–183. ACM Press, 2005. (Cited on page 33.)

[MEM] *Memex project - http://en.wikipedia.org/wiki/Memex*. (Cited on page 39.)

[MIT] *Massachusetts Institute of Technology (MIT) - http://www.mit.edu/*. (Cited on page 41.)

[MOS] *Microsoft Office SharePoint Server - http://facetedsearch.codeplex.com/*. (Cited on page 37.)

[Mukhopadhyay 1999] Sugata Mukhopadhyay and Brian Smith. *Passive capture and structuring of lectures*. In Proceedings of the seventh ACM international conference on Multimedia (Part 1), pages 477–487, 1999. (Cited on page 71.)

[Munir 2007] K. Munir, M. Odeh, R. McClatchey, S. Khan and I. Habib. *Semantic Information Retrieval from Distributed Heterogeneous Data Sources*. ArXiv e-prints, 2007. (Cited on page 38.)

[MUV] *MUVINO - multimedia video annotation tool - http://vitooki.sourceforge.net/components/muvino/code/*. (Cited on page 25.)

[Najood Al-Ghamdi 2010] Mostafa Saleh Najood Al-Ghamdi and Fathy Eassa. *Ontology-Based Query in Heterogeneous and Distributed Data Sources*. Asian Journal of Information Technology, vol. 9, pages 183–195, 2010. (Cited on page 38.)

[Nan 2009] Xiaoming Nan and al. *A novel framework for semantic-based video retrieval*. 2009. (Cited on page 44.)

[NBC] (Cited on page 11.)

[Nejdl 2002] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjorn Naeve, Mikael Nilsson, Matthias Palmér and Tore Risch. *EDUTELLA: a P2P networking infrastructure based on RDF*. In Proceedings of the 11th international conference on World Wide Web, WWW '02, pages 604–615. ACM, 2002. (Cited on page 48.)

[ONT] *Expert Glossary - ontology definition - http://www.expertglossary.com/knowledge-management/definitions/0*. (Cited on page 18.)

[organisation for standardisation 2004] International organisation for standardisation. *MPEG7 overview*. October 2004. (Cited on page 26.)

[OWL] *Web Ontology Language (OWL) - http://www.w3.org/2007/OWL*. (Cited on page 18.)

[P2P] (Cited on page 51.)

[Perugini 2010] Saverio Perugini. *Supporting multiple paths to objects in information hierarchies: Faceted classification, faceted search, and symbolic links*. Information Processing and Management, vol. 46, no. 1, pages 22–43, 2010. (Cited on page 29.)

[Petridis 2005] Kosmas Petridis and al. *M-ontomat-annotizer, A tool for semantic annotation of images and videos for multimedia analysis and retrieval*. 2005. (Cited on page 26.)

[PHA] *PHAROS: Platform for searcHing of Audiovisual Resources across Online Spaces*. http://www.pharos-audiovisual-search.eu/. (Cited on page 44.)

[POL] *Polysema - Multimedia application supported by semantics - http://polysema.di.uoa.gr/*. (Cited on page 27.)

[Polowinski 2009] Jan Polowinski. *Widgets for Faceted Browsing*. In Michael Smith and Gavriel Salvendy, editeurs, Human Interface and the Management of Information. Designing Information Environments, volume 5617 of *Lecture Notes in Computer Science*, pages 601–610. Springer Berlin / Heidelberg, 2009. (Cited on page 29.)

[PRE] *Prefuse library - http://prefuse.org*. (Cited on page 118.)

[Pridmore-Brown 2010] Philip Pridmore-Brown. *An Inventory of Semantic Models (Ontologies) for Use When Managing Enterprise Tacit Knowledge*. 2010. (Cited on pages 17 and 19.)

[PRO a] *Processing* - *http://www.processing.org*. (Cited on page 118.)

[PRO b] *Protege - ontology editor and knowledge acquisition system* - *http://protege.stanford.edu/*. (Cited on page 21.)

[PYT] *The Python Python Programming Language* - *http://python.org/*. (Cited on page 107.)

[Quan 2003] Dennis Quan, David Huynh and David R. Karger. *Haystack: A Platform for Authoring End User Semantic Web Applications*. In In ISWC, pages 738–753, 2003. (Cited on page 33.)

[Ramzan 2010] Naeem Ramzan, Charalampos Patrikakis, Qianni Zhang and Ebroul Izquierdo. *Analysing multimedia content in social networking environments*. In Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access, SAPMIA '10, pages 73–76, New York, NY, USA, 2010. ACM. (Cited on pages 4 and 43.)

[RAP] *Raphael* - *http://raphaeljs.com*. (Cited on page 118.)

[Rasiwasia 2010] Nikhil Rasiwasia, Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert R.G. Lanckriet, Roger Levy and Nuno Vasconcelos. *A new approach to cross-modal multimedia retrieval*. In Proceedings of the international conference on Multimedia, MM '10, pages 251–260, New York, NY, USA, 2010. ACM. (Cited on page 38.)

[RDF] *Resource Description Framework (RDF)* - *http://www.w3.org/RDF/*. (Cited on page 18.)

[Revertera 2011] Jean Revertera, Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *SMAC: Smart Multimedia Archiving for Conferences*. In AWIC, pages 143–152, 2011. (Cited on page 6.)

[Sabol 2008] Vedran Sabol, Wolfgang Kienreich and Michael Granitzer. *Visualisation Techniques for Analysis and Exploration of Multimedia Data*. In Michael Granitzer, Mathias Lux and Marc Spaniol, editeurs, Multimedia Semantics - The Role of Metadata, volume 101 of *Studies in Computational Intelligence*, pages 219–238. Springer Berlin / Heidelberg, 2008. (Cited on pages 32 and 42.)

[Saleh] Mostafa E. Saleh. *Semantic Query in Heterogeneous Web Data Sources*. I. J. Comput. Appl. (Cited on page 38.)

[SAP] *Search in audio visual content using P2P IR (SAPIR)* - *http://www.sapir.eu/tech//underscore//obj.html*. (Cited on page 48.)

[Sattler 2005] Kai-Uwe Sattler, Ingolf Geist and Eike Schallehn. *Concept-based querying in mediator systems*. The VLDB Journal, vol. 14, pages 97–111, 2005. (Cited on page 39.)

[Sauermann 2003] Leo Sauermann. The gnowsis – using semantic web technologies to build a semantic desktop. Master's thesis, 2003. (Cited on page 41.)

[Sauermann 2005] Leo Sauermann, Ansgar Bernardi and Andreas Dengel. *Overview and Outlook on the Semantic Desktop.* In In Proc. of Semantic Desktop Workshop at the ISWC, 2005. (Cited on page 41.)

[Schmidt 2010] Matthias Schmidt, Jan Polowinski, Jendrik Johannes and Miguel Fernandez. *An Integrated Facet-Based Library for Arbitrary Software Components.* In Thomas Kühne, Bran Selic, Marie-Pierre Gervais and Francois Terrier, editeurs, Modelling Foundations and Applications, volume 6138 of *Lecture Notes in Computer Science*, pages 261–276. Springer Berlin / Heidelberg, 2010. (Cited on page 29.)

[Schroeter 2003] Ronald Schroeter, Jane Hunter and Douglas Kosovic. *Vannotea - a collaborative video indexing, annotation and discussion system for broadband networks.* In In K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation, 2003. (Cited on pages 25 and 26.)

[SEM] *SEMantic Peer Layer InfrastruCturE (SEMPLICE)* `http://semantics.eng.it/semplice/`. (Cited on page 48.)

[Sidiropoulos 2005] Antonis Sidiropoulos and Yannis Manolopoulos. *A New Perspective to Automatically Rank Scientific Conferences Using Digital Libraries.* In In Information Processing and Management 41 (2005) 289Z12, pages 289–312, 2005. (Cited on page 11.)

[SIR] *Semantic Integration Re-flecting User-specific semantic (SIRUP)* - `http://www.ifi.uzh.ch/dbtg/Projects/SIRUP/`. (Cited on page 41.)

[Smeulders 2000] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta and Ramesh Jain. *Content-Based Image Retrieval at the End of the Early Years.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, pages 1349–1380, 2000. (Cited on page 3.)

[Snoek 2005] Cees G. M. Snoek and Marcel Worring. *Multimodal Video Indexing: A Review of the State-of-the-art.* Multimedia Tools Appl., vol. 25, pages 5–35, 2005. (Cited on page 62.)

[Sokhn 2009a] Maria Sokhn, Francesco Carrino, Elena Mugellini, Omar Abou Khaled and Ahmed Serhrouchni. *Conference knowledge modeling for conference-video-recordings querying visualization.* In MEDES, pages 226–232. ACM, 2009. (Cited on page 6.)

[Sokhn 2009b] Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Knowledge management framework for conference video-recording retrieval.* 2009. (Cited on page 6.)

[Sokhn 2009c] Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Poster: Knowledge based video-recordings visualization.* 2009. (Cited on page 6.)

[Sokhn 2009d] Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Knowledge modeling for enhanced information retrieval and visualization.* In Atlantic Web Intelligence Conference (AWIC), 2009. (Cited on page 6.)

[Sokhn 2010a] Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Poster: Conference advanced level information management and retrieval.* 2010. (Cited on page 6.)

[Sokhn 2010b] Maria Sokhn, Elena Mugellini, Omar Abou Khaled and Ahmed Serhrouchni. *Conference knowledge modelling for conference-video-recordings querying and visualization.* Multimedia Processing and Technologies, vol. 1, pages 95–111, 2010. (Cited on page 6.)

[Sokhn 2011a] Maria Sokhn, Elena Mugellini, Omar Abou Khaled and Ahmed Serhrouchni. *End-to-End Adaptive Framework for Multimedia Information Retrieval.* vol. 6649, pages 197–206, 2011. (Cited on page 6.)

[Sokhn 2011b] Maria Sokhn, Elena Mugellini and Omar Abou Khaled. *Querying over heterogeneous and distributed data sources.* In Atlantic Web Intelligence Conference (AWIC), pages 29–38, 2011. (Cited on page 6.)

[SPA] *SPARQL Protocol and RDF Query Language (SPARQL) - http://www.sparql.org/.* (Cited on page 33.)

[SPH] *sphinx4 - speech recognizer - http://cmusphinx.sourceforge.net/sphinx4/.* (Cited on page 76.)

[Stoica 2003] I. Stoica, R. Morris, D. Liben-Nowell, D.R. Karger, M.F. Kaashoek, F. Dabek and H. Balakrishnan. *Chord: a scalable peer-to-peer lookup protocol for Internet applications.* Networking, IEEE/ACM Transactions on, vol. 11, pages 17 – 32, 2003. (Cited on pages 49 and 51.)

[Su 2010a] Addison Y. S. Su, Stephen J. H. Yang, Wu-Yuin Hwang and Jia Zhang. *A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments.* Comput. Educ., vol. 55, pages 752–766, 2010. (Cited on page 26.)

[Su 2010b] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh and Vincent S. Tseng. *Effective content-based video retrieval using pattern-indexing and matching techniques.* Expert Syst. Appl., vol. 37, pages 5068–5085, 2010. (Cited on page 43.)

[Sundara 2010] Seema Sundara, Medha Atre, Vladimir Kolovski, Souripriya Das, Zhe Wu, Eugene Inseok Chong and Jagannathan Srinivasan. *Visualizing large-scale RDF data using Subsets, Summaries, and Sampling in Oracle.*

Data Engineering, International Conference on, vol. 0, pages 1048–1059, 2010. (Cited on page 33.)

[Sure 2002] York Sure, Michael Erdmann, Juergen Angele, Steffen Staab, Rudi Studer and Dirk Wenke. *Ontoedit: Collaborative ontology development for the semantic web.* pages 221–235. Springer, 2002. (Cited on page 21.)

[SWC] *Semantic Web Conference (SWC) ontology* - *http://data.semanticweb.org/ns/swc/swc_2009-05-09.html*. (Cited on pages 23 and 64.)

[SWO] *SWOOKI* *http://www.loria.fr/~skaf/pmwiki/pmwiki.php/Main/SWOOKI*. (Cited on page 48.)

[SWR] *Semantic Web for Research Communities (SWRC)* - *http://ontoware.org/swrc/*. (Cited on pages 22 and 64.)

[T. E. El-Diraby 2005] K. F. Kashif T. E. El-Diraby. *ODistributed Ontology Architecture for Knowledge Management in Highway Construction.* volume 131, pages 591–603. ACM, 2005. (Cited on page 19.)

[Tallis 2003] M. Tallis. *Semantic word processing for content authors.* In Proceedings of the Knowledge Markup & Semantic Annotation Workshop, Florida, USA, numéro Semannot. Citeseer, 2003. (Cited on page 24.)

[Thomas M. 2007] Jean-Marc O. Thomas M. Alain B. *Multimedia scenario extraction and content indexing.* pages 204–211. CBMI, 2007. (Cited on page 108.)

[Tjondronegoro 2005a] Dian Tjondronegoro, Amanda Spink and Bernard J. Jansen. *Search Engine Users: Internet searchers are confident, satisfied and trusting - but they are also unaware and naive.* 2005. (Cited on page 29.)

[Tjondronegoro 2005b] Dian W. Tjondronegoro. *PhD Thesis: "Content-based Video Indexing for Sports Applications using Multi-modal approach".* PhD thesis, Deakin University, 2005. (Cited on pages 17 and 24.)

[Tjondronegoro 2008] Dian Tjondronegoro and Amanda Spink. *Web search engine multimedia functionality.* Inf. Process. Manage., vol. 44, pages 340–357, 2008. (Cited on page 28.)

[Tjondronegoro 2009] Dian Tjondronegoro, Amanda Spink and Bernard J. Jansen. *A study and comparison of multimedia Web searching: 1997-2006.* JASIST, pages 1756–1768, 2009. (Cited on pages 28, 29, 37 and 44.)

[tram Dang-ngoc 2005] Tuyet tram Dang-ngoc, Clement Jamard and Nicolas Travers. *XLive: An XML Light Integration Virtual Engine.* In In Proc. of BDA, 2005. (Cited on page 39.)

[Tvarozek 2008] Michal Tvarozek and Maria Bielikova. *Collaborative multi-paradigm exploratory search.* In Proceedings of the hypertext 2008 workshop on Collaboration and collective intelligence, WebScience '08, pages 29–33. ACM, 2008. (Cited on page 31.)

[Tzitzikas 2005] Yannis Tzitzikas, Nicolas Spyratos and Panos Constantopoulos. *Mediators over taxonomy-based information sources.* The VLDB Journal, vol. 14, pages 112–136, 2005. (Cited on page 38.)

[Umar 2009] Amjad Umar and Adalberto Zordan. *Enterprise Ontologies for Planning and Integration of Business: A Pragmatic Approach.* IEEE Transactions on Engineering Management, vol. 56, no. 2, pages 352–371, 2009. (Cited on page 18.)

[UZH] *University of Zurich* - http://www.uzh.ch/. (Cited on page 41.)

[Vembu 2006] Shankar Vembu, Malte Kiesel, Michael Sintek and Stephan Baumann. *Towards bridging the semantic gap in multimedia annotation and retrieval.* In In: Proceedings of the 1st International Workshop on Semantic Web Annotations for Multimedia at the 15th International World Wide Web Conference, 2006. (Cited on page 24.)

[VID a] *VideoAnnEx - IBM MPEG7 Annotation Tool* - http://www.alphaworks.ibm.com/tech/videoannex. (Cited on page 25.)

[VID b] *Videolectures portal* - http://www.rkbexplorer.com/. (Cited on page 34.)

[VIE] *Vienna university of technology* - http://www.tuwien.ac.at/tu_vienna/. (Cited on page 39.)

[VIK] *An advanced software framework for enabling the integrated development of semantic-based Information, Content, and Knowledge (ICK) management systems.* FP6 Framework. http://www.vikef.net/. (Cited on pages 15 and 44.)

[VIS] *DBLPVis - a visual feature for the DBLP library* - http://www.faviki.com/website/dblpvis.uni-trier.de. (Cited on page 34.)

[Visualization 2005] Adapting Graph Visualization, Flavius Frasincar, Ru Telea and Geert jan Houben. *Adapting graph visualization techniques for the visualization of RDF data.* In of RDF data, Visualizing the Semantic Web, 2006, pages 154–171, 2005. (Cited on page 33.)

[VOC] *A vocabulary for describing relationships between people* - http://vocab.org/relationship/. (Cited on page 21.)

[W3C] *World Wide Web (W3C)* - http://www.w3.org/. (Cited on page 18.)

[Wang 2003] Feng Wang, Chong-Wah Ngo and Ting-Chuen Pong. *Synchronization of lecture videos and electronic slides by video text analysis*. In Proceedings of the eleventh ACM international conference on Multimedia, pages 315–318, 2003. (Cited on page 71.)

[Wang 2011] Lei Wang, Dawei Song and Eyad Elyan. *Video retrieval based on words-of-interest selection*. In Proceedings of the 33rd European conference on Advances in information retrieval, ECIR'11, pages 687–690, 2011. (Cited on page 43.)

[WEB] *Web Ontology Developement Environement (WebODE) -* `http://mayor2.dia.fi.upm.es/oeg-upm/`. (Cited on page 21.)

[Weiand 2008] K. Weiand, F. Bry and T. Furche. *Reasoning & Querying-State of the Art: Keyword-based querying for XML and RDF*. In Institute for Informatics, University of Munich. Institute for Informatics, University of Munich, 2008. (Cited on page 38.)

[Weiland 2008] Markus Weiland and Raimund Dachselt. *Facet folders: flexible filter hierarchies with faceted metadata*. In CHI '08: CHI '08 extended abstracts on Human factors in computing systems, pages 3735–3740, New York, NY, USA, 2008. ACM. (Cited on page 31.)

[Wiederhold 1992] Gio Wiederhold. *Mediators in the architecture of future information systems*. IEEE COMPUTER, vol. 25, no. 3, pages 38–49, 1992. (Cited on page 38.)

[WIK] *Wikipedia -* `http://www.wikipedia.org/`. (Cited on page 18.)

[Wilkins 2007] Peter Wilkins, Tomasz Adamek, Daragh Byrne, Gareth J. F. Jones, Hyowon Lee, Gordon Keenan, Kevin McGuinness, Noel E. O'Connor, Alan F. Smeaton, Alia Amin, Zeljko Obrenovic, Rachid Benmokhtar, Eric Galmar, Benoit Huet, Slim Essid, Rémi Landais, Felicien Vallet, Georgios Papadopoulos, Stefanos Vrochidis, Vasileios Mezaris, Ioannis Kompatsiaris, Evaggelos Spyrou, Yannis S. Avrithis, Roland Mörzinger, Peter Schallauer, Werner Bailer, Tomas Piatrik, Krishna Chandramouli, Ebroul Izquierdo, Martin Haller, Lutz Goldmann, Amjad Samour, Andreas Cobet, Thomas Sikora and Pavel Praks. *K-Space at TRECVID 2007*. In Paul Over, George Awad, Wessel Kraaij and Alan F. Smeaton, editeurs, TRECVID. National Institute of Standards and Technology (NIST), 2007. (Cited on page 25.)

[Wilson 2008] Max L. Wilson and m.c. schraefel. *A longitudinal study of exploratory and keyword search*. In Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries, JCDL '08, pages 52–56. ACM, 2008. (Cited on page 31.)

[WOR] *Wordnet - a lexical database for the English language -* `http://www.cogsci.princeton.edu/wn/`. (Cited on page 19.)

[XML] *eXtensible Markup Language (XML)* - *http://www.w3.org/XML/*. (Cited on page 20.)

[XQU] *XML Query (XQuery)* - *http://www.w3.org/XML/Query/*. (Cited on page 83.)

[Xu 2009] Cheng Zhi Xu, Hui Chuan Wu, Bo Xiao, Philip C. Y. Sheu and Shu-Ching Chen. *A Hierarchical Video Annotation System.* In Proceedings of the 2009 IEEE International Conference on Semantic Computing, ICSC '09, pages 670–675. IEEE Computer Society, 2009. (Cited on pages 24 and 25.)

[Yeh 2010] Ching-Long Yeh, Chun-Fu Chang and Kuo-Chung Lin. *Two-Level Chord-Ring for Ontology-based Distributed RDF Store.* In Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems, CISIS '10, pages 283–290. IEEE Computer Society, 2010. (Cited on page 51.)

[Yuan 2007] Jinhui Yuan, Huiyi Wang, Lan Xiao, Wujie Zheng, Jianmin Li, Fuzong Lin and Bo Zhang. *A formal study of shot boundary detection.* In IEEE transactions on circuits and systems for video technology, pages 168–186, 2007. (Cited on page 71.)

[Z. 2008] Na Z. and al. *An Integrated and Interactive Video Retrieval Framework with Hierarchical Learning Models and Semantic Clustering Strategy.* 2008. (Cited on page 44.)

[Zhang 2002] Hong-Jiang Zhang. *Content-based video analysis, retrieval, and browsing.* 2002. (Cited on page 24.)

[Zhang 2008] Ji Zhang. *A review of ontology and web service composition.* 2008. (Cited on pages 18 and 20.)

Part VII

Appendix

Appendix A

Annex

HELO model extracts

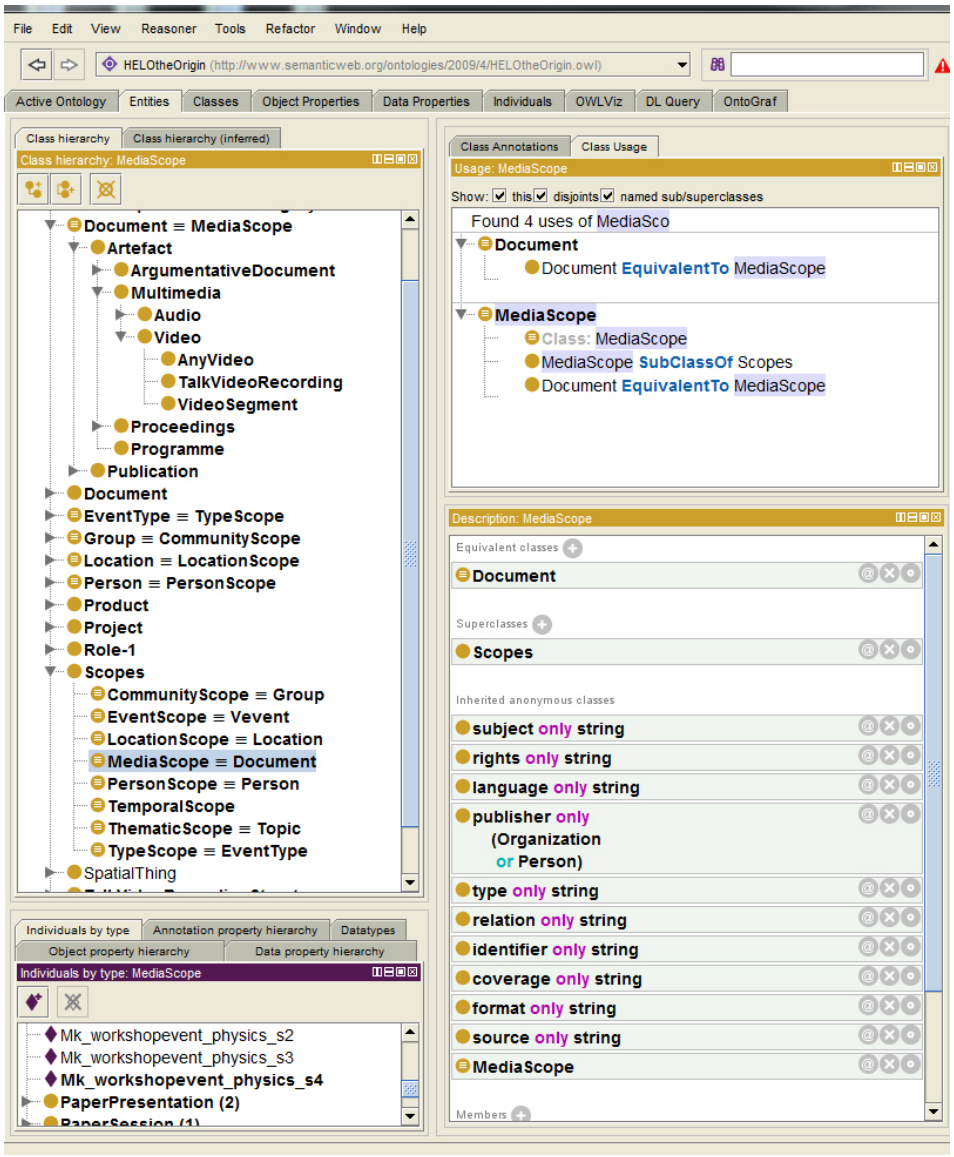


Figure A.1: HELO - an extract - I

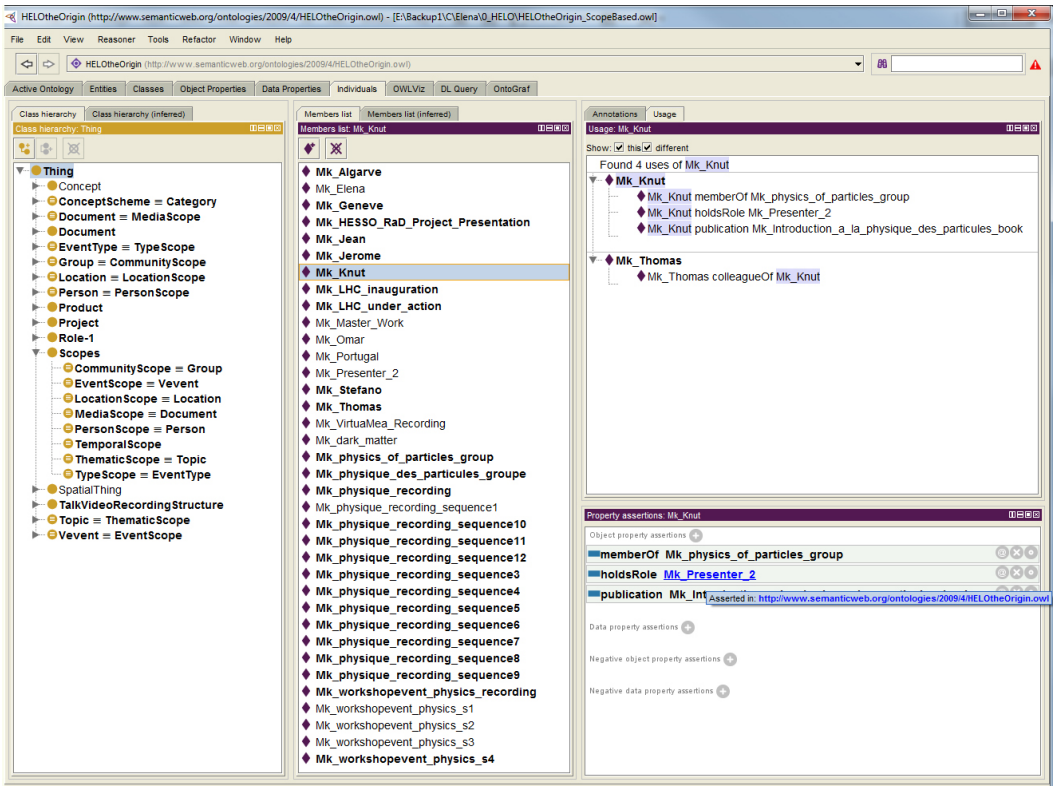


Figure A.2: HELO - an extract - II

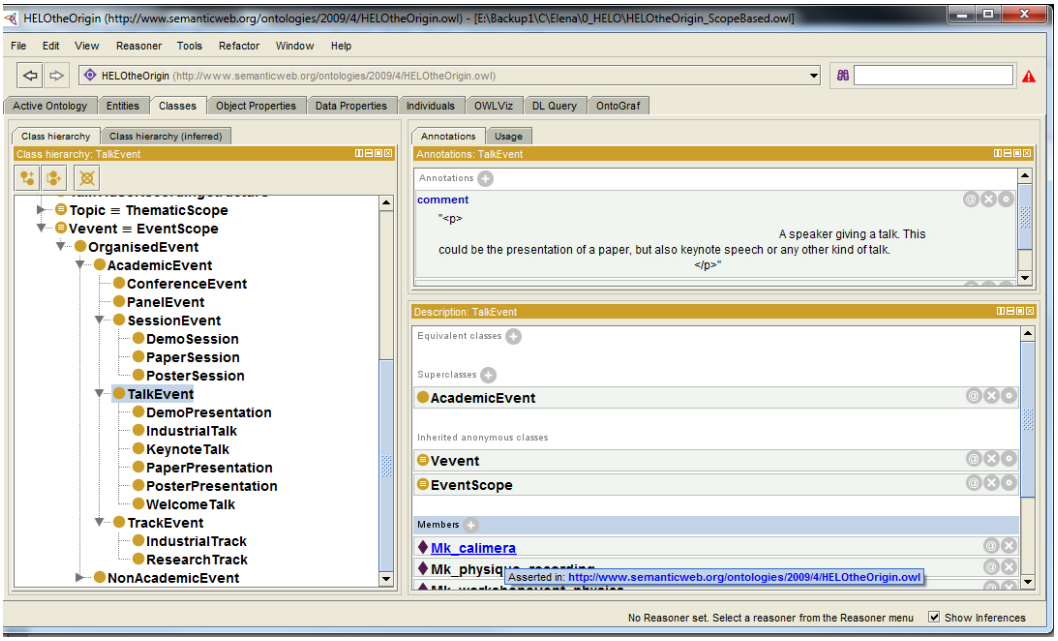


Figure A.3: HELO - an extract - III

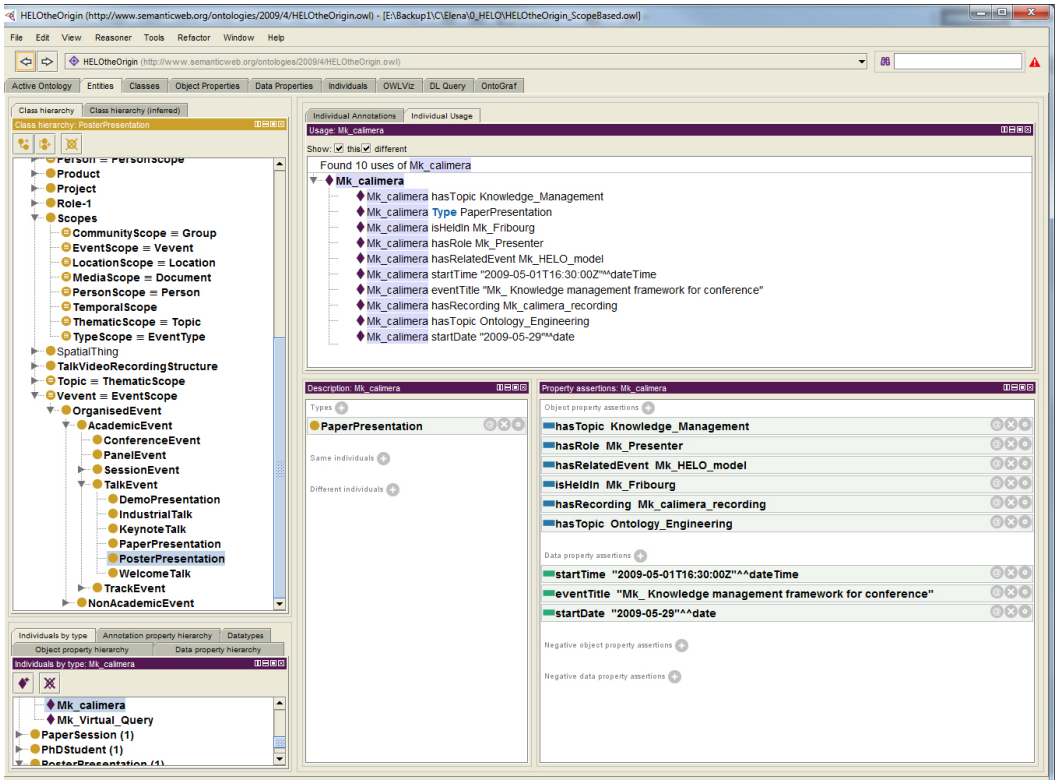


Figure A.4: HELO - an extract - IV

Appendix B

Glossary

ACM: Association for Computing Machinery
Adobe Flash: A multimedia platform used to add animation, video, and interactivity to web pages
Advene: Annotate Digital Video, Exchange on the Net
AKT: Advanced Knowledge Technologies
Annomation: A tool for semantically annotating motion video
Anvil: A video annotation tool
API: Application Programming Interface
Arnetminer: Expertise oriented search system
Ask-GraphView: A large scale graph visualization system
AWIC2011: Antlatic Web Intelligence Conference 2011
BibTEX: A reference management software for formatting lists of references
CALIMERA: Conference Advanced Level Information ManagemEnt and RetrievAl
Caliph & Emir: MPEG7 based Java prototypes for digital photo and image annotation and retrieval
CALISEMA: Semantic based annotation tool integrated into CALIMERA and based on POLYSEMA
CBMR: Content Based Multimedia Retrieval
CDS: CERN Document Server
CERN: Centre Europeen de Recherche Nucleaire
Chord: A protocol and algorithm for a peer-to-peer distributed hash table
COALA: Content-Oriented Audiovisual Library Access
Conference: In our work, we define the word "conference" as follows: a conference, is any type of meeting, happening at a determinable time and place, with at least one speaker who intend to inform, explain, or talk about one or several topics
ConKMeL: A contextual knowledge management framework to support multimedia e-learning
CPU: Central Processing Unit
CREAM: CREAting Metadata for the semantic web
Crescendo: A comprehensive speech recognition solution
DAML+OIL: DARPA Agent Markup Language+Ontology Inference Layer
DARPA: Defense Advanced Research Projects Agency
DBLP: Digital Bibliography and Library Project
DBLPViz: A visual feature for the DBLP library
Dbpedia: A project aiming to extract structured content from the information created as part of the wikipedia project
DCT: Discrete Cosine Transform
DESY: Deutsches Elektronen Synchrotron
DHT: Distributed Hash Table

Dragon: Naturally Speaking A speech recognition software package
eBiquity: A research group from the university of Maryland, Baltimore County
EIAFR: Ecole d'Ingenieurs et d'Architecte de Fribourg
ELAN: Educational LANguage
EPFL: Ecole Polytechnique Federale de Lausanne
ESWC: European Semantic Web Conference
EXMARaLDA: EXtensible MARkup Language for Discourse Annotation
Flare: An ActionScript library for creating visualizations that run in the adobe flash player
FNAL: Fermi National Accelerator Laboratory
FOAF: Friend Of A Friend
FS: Following Sequence
GViz: Geo-Visualization: a viisualization middleware for e-science
Haystack: A semantic web browser developed by the haystack project at the MIT laboratory
HCI: Human Computer Interaction
HELO: High-level modEL Ontology - an ontology model for conference
IBM: International Business Machines
ICK: Information, Content, and Knowledge
iFinder: An MPEG7-based retrieval system for distributed multimedia content
INDICO: A tool for the conferences, workshops and meetings management
InfoViz: A tools for creating Interactive data visualizations for the web
INVENIO: A software to run a digital library or document repository on the web
IR: Information Retrieval
IsaViz: A visual environment for browsing and authoring RDF models
ISWC: Internatioal Semantic Web Conference
Jena: A java framework for building semantic web applications
JOANNEUM RESEARCH: A professional innovation and technology provider
JXTA: A language and platform independent protocol for P2P networking
JXTACh: A chord-based JXTA version
K-Space: A framework for semi-automatic, semantic annotation of multimedia content
KVC: Knowledge Value Chain
LHC: Large Hadron Collider
Matterhorn: A platform to support the management of educational audio and video content
MDS: Multimedia Description Scheme
MEDES: Management of Emergent Digital EcoSystems
MISG: Multimedia Information System Group
MIT: Massachusetts Institute of Technology
M-OntoMat-Annotizer: Multimedia OntoMat-Annotizer - OntoMat-Annotizer: An interactive webpage annotation tool
MOSS: Microsoft Office SharePoint Server **MPEG7:** Moving Picture Experts Group 7 - a multimedia content description standard
mSpace: A multi faceted column based client for exploring large data sets
MuViNo: A tool for creating MPEG7 metadata of videos
NAVIR: An ontology based tool for multimedia browsing and visualization
NEPOMUK: Networked Environment for Personalized, Ontology-based Management of Unified Knowledge
OAI-PMH: Open Archives Initiative-Protocol for Metadata Harvesting

OCR: Optical Character Recognition
OntoEdit: OntoEdit is a development environment for design, adaptation and import of knowledge models for application systems
OWL: Ontology Web Language
P2P: Peer to Peer
PAMS: Production, Advertising, Merchandising Service
PHAROS: A european research project on audiovisual search
POLYSEMA: An MPEG7 Video Annotator
Processing: An programming language and environment for creating images, animations, and interactions
Protege2000: An integrated software tool to develop knowledge-based systems
PS: Previous Sequence
Raphael: A javascript library for vector graphics implementation on the web
RCSO: Reseau de Competences de Suisse occidentale
RDF: Resource Description Framework
RDF: Gravity A tool for visualising RDF/OWL graphs/ontologies
RDFS: Resource Description Framework Schema
ReSIST: Resilient Computing Curriculum and Courseware
RKBExplorer: ReSIST Knowledge Base Explorer
SAPIR: Search in audio visual content using P2P IR
SemanticLIFE: A virtual query system
SEMPLICE: SEMantic Peer Layer InfrastruCturE
SIFT: Scale Invariant Feature Transform
SINGAPORE: SINGle Access POint for heterogeneous data Reposiotries
SIRUP: Semantic Integration Re-flecting User-specific semantic
SLAC: Stanford Linear Accelerator Center
SMAC: Smart Multimedia Archive for Conference
SPARQL: SPARQL Protocol and RDF Query Language
Sphinx4: A speech recognition system
SQL: Structured Query Language
SWC: Semantic Web Conference
TSIMMIS: The Stanford-IBM Manager of Multiple Information Sources
URI: Uniform Resource Identifier
Vannotea: A collaborative video indexing, annotation and discussion system
VideoAnnEx: A tool for annotating video sequences with MPEG7 metadata
VIKEF: Virtual Information and Knowledge Environment Framework
Virtual-Q: Virual Query
VQE: Virtual Query Engine
W3C: World Wide Web Consortium
WebODE: Web Ontology Developement Environement - an extensible ontology-engineering suite based on an application server
Welkin: Welkin is a graph-based RDF visualizer
Wordnet: A lexical database of English
XLive: An XML Light Integration Virtual Engine
XML: eXtensible Markup Language
XQuery: XML query

Appendix C

Curriculum Vitae