



## DISCOVERING SEMANTICS IN MULTIMEDIA CONTENT USING WIKIPEDIA

ANGELA FOGAROLLI AND MARCO RONCHETTI\*

**Abstract.** Semantic-based information retrieval is an area of ongoing work. In this paper we present a solution for giving semantic support to multimedia content information retrieval in an e-Learning environment where very often a large number of multimedia objects and information sources are used in combination. Semantic support is given through intelligent use of Wikipedia in combination with statistical Information Extraction techniques.

**Key words:** content retrieval and filtering; search over semi-structural Web sources, multimedia, Wikipedia, e-Learning

**1. Introduction.** Nowadays, organizations have to deal with information overloading. They need a way to organize and store their content and being able to easily retrieve it when necessary. Our objective is to provide a system for indexing and retrieving content based on the semantic provide by Wikipedia. Retrieving the desired content can be difficult due to the the high specifically of terms in a search task.

In our work, we are addressing the problem of accessing different kinds of unstructured or semi-structured information sources taking advantages of the semantic provided by public available resources such as Wikipedia. Furthermore using the approach we will describe in section 4 we would like to automatize the task of annotating a corpus and discover relations between annotations. Next we will use annotation in combination with textual information retrieval for determining the search context and based on it we will be able to give search suggestions and perform query expansion. Using annotation in information retrieval is not a new idea [6, 4] even in combination with ontologies [3], it has been widely used in video and image retrieval generating also a social phenomena like folksonomy [15, 13]. What is new is the use of domain independent public available semantic to automatically describe content in different kind of media.

We are applying our approach in the e-Learning context, specifically enhanced streaming video lectures (see [8, 5]) because of the peculiarity in this scenario of combining different kinds of unstructured or semi-structured sources of information. E-Learning presents many problematics in common with the business scenario in terms of content classification for its amount of information to classify and for the different contexts where a specific information can be relevant. Our target repository collects different kinds of media (video, audio, presentation slides, text documents), which can be searched and presented in combination. For each recorded event (e.g. lecture, seminar, talk, meeting) we provide not only the video but also related materials, which can consist of presentation slides, documents or Web sites the speaker points to. All the resources are temporally synchronized with the video.

An example of how a multimedia presentation of this kind looks like can be found in figure 1.1; the video with the speaker appears together with presentation slides or additional notes. Video and slides are synchronized and can be navigated by means of a temporal bar or by slide titles.

We can summarize the following five state of the art approaches to multimedia indexing and navigation:

1. Use of metadata to browse keyframes.
2. Use text from speech, using transcript-based search.
3. Matching keyframes vs. querying of images. Keyframes extracted as shot representatives are used for retrieval. It requires user to locate images/other keyframes, from browsing or other search.
4. Use of semantic features. They are based upon pre-processing video or keyframes to detect features. Features can be related to ontologies.
5. Use video/image objects as queries.

We concentrate on pt. (2) and partially on pt. (3), we use the text-from-speech technique combined with a textual analysis of the speech and the event related material using Wikipedia instead of ontologies.

In Wikipedia, the concept of class and instance are not separated as in the ontological sense, due to the fact that it is not constrained to a formal model, for the reason of which it is not possible to formalize reasoning on the Wikipedia content directly.

The use of Wikipedia url as suggested in[10] for concept identification could guarantee interoperability between domain ontologies, while the extensive ongoing research effort for extracting an ontological view from

\*University of Trento, Dept. of Information and Communication Tech., Via Sommarive 14, 38050 Trento, Italy  
{angela.fogarolli, marco.ronchetti}@unitn.it

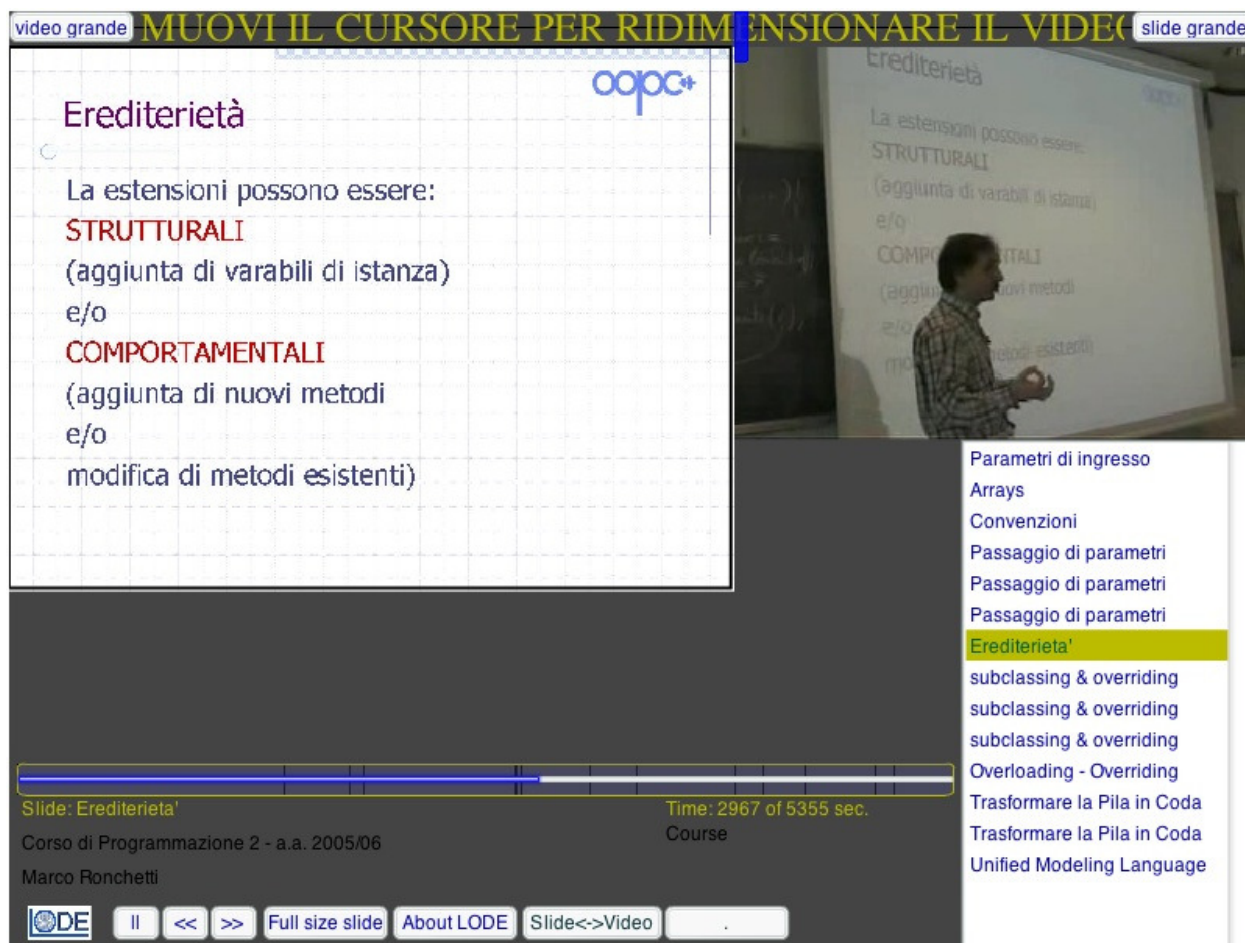


FIG. 1.1. An example of multimedia presentation

Wikipedia could soon lead to the creation of an ontological view based on Wikipedia which could be the reference for many domains.

The combination of information extracted from video and related material gives a complete picture of an event, since in the real world the sum of *all* the media used by a speaker is meant to fully describe the event's topics to facilitate knowledge transfer to the audience.

In this paper we report about how we provide semantic support and unsupervised annotation of multimedia material based on information extracted from Wikipedia, rather than the usage of Semantic Web technologies (specifically without ontologies). Our approach is domain independent, and in theory it could also be applied to different use cases where there is a need for clustering or annotation of a corpus.

The structure of this paper is organized as follows: in the next section we describe the context and the motivation of our work; section 4 gives an overview of our approach. In section 5 we apply the approach described in the previous section to our use case. Finally, we discuss the directions we are planning to take regarding further work.

**2. Semantics in the Web.** In the Web, some collections of data containing semantic annotations (e.g. UniProt: <http://www.ebi.ac.uk/swissprot/index.html>, Ecademy: <http://www.ecademy.com>) are now available and there is a trend to semantically enable more and more Web content. Even though this trend is perceivable, there is still a huge amount of material on which these technologies have not been applied. One limiting factor for a faster adoption of Semantic Web technologies, is the difficulty to find ready-to-use conceptualizations for annotating existing material and making it Semantic Web compatible.

We explored the possibility of using Semantic Web ontologies for annotating multimedia material and for discovering and presenting to the user relations between the searched topics and other topics, based on the

relationships between entities in one or more domain ontologies. We experienced difficulties in finding ontologies which cover a variety of domains, since e-Lectures can cover an unpredictable amount of domains (e.g. computer science, history, meteorology, geography, math...). In addition, the terms expressed in e-Lectures are usually *individuals* of an ontology (e.g. the term 'Collection' in a Java Programming class could be modelled as an instance of a data container class in a Java Programming ontology) and finding populated ontologies with a wide coverage of individuals to date is a big challenge, and usually requires the involvement of a knowledge engineer.

Our requirement was to find a broad, domain-independent collection of individual terms (as opposed to concepts) which are connected by relations. To the best of our knowledge, the most complete collection of this kind is Wikipedia. Wikipedia is a freely available encyclopedia which is constantly growing in size and in fame thanks to the copyleft license that allows the content to be copied, modified and redistributed as long as there is an acknowledgment of the author and the new content is published under the same license (see <http://www.wikipedia.org>). Wikipedia contains a classification of topics, organized with an hierarchy of categories and with relationships between elements. The advantage of using it is that the social collaborative network around it makes its content always up to date and it covers in details a huge amount of topics in different domains and languages. In addition it also takes into account the different possible meanings of a term through a disambiguation page.

What we can extract using Wikipedia are the relationships between topics. According to Obrst's definitions in [11], Wikipedia not only offers *weak semantic* information, such as parent-child relationships, but it also contains lexicographic relationships that—once the domain of interest is determined—can offer *medium semantic*. In Wikipedia we do not have *strong semantic*, i. e. we can not describe real-world relationships such as "a car has a minimum of four wheels" as with the usage of an ontology. We can only deduce that concepts are connected without knowing how; we can tell that one concept in one category is related to other concepts which are linked in the description of the concept itself.

In Wikipedia, the concept of class and instance are not separated as in the ontological sense, due to the fact that it is not constrained to a formal model, for the reason of which it is not possible to formalize reasoning on the Wikipedia content directly. There are projects (see section 3) that try to embed semantic inside Wikipedia extending the Wiki software used to write Wikipedia pages [16], and some others (e.g. [www.dbpedia.org](http://www.dbpedia.org) [2]) which provide an RDF representation of Wikipedia, to make its content machine-interpretable.

We use Wikipedia as a taxonomy to obtain lexicographic relationships and in combination with statistical information extraction we can deduce related concepts to the terms extracted from our corpus. In addition, since our corpus covers a representation of a part of the real world we also use the corpus itself as "training data" for domain disambiguation in Wikipedia.

There is a lot of work about extracting semantics (some is reported in section 3) from Wikipedia content to build an ontological representation. We are therefore confident that even though for now we can extract information without a high semantic value in the feature, at the light and with the combination of other research effort in the area we will be able to increase the power of our approach in terms of flexibility, extension and accuracy.

**3. Related Work.** Wikipedia contains a vast amount of information, therefore there have been mainly two approaches for exploring its content and make it machine readable. The first approach consists in embedding semantic notations in its content [16, 7]; while the other one deals with information extraction based on the understanding of how the Wikipedia content is structured: [1, 14, 17, 12, 19].

The SemanticWikipedia project [16] is an initiative that invites Wikipedia authors to add semantic tags to their articles in order to make them machine interpretable. The wiki software behind Wikipedia (MediaWiki [7]), itself enables authors to represent structured information in an attribute-value notation, which is rendered inside a wiki page by means of an associated template.

The second main stream of Wikipedia related work is on automatically extract knowledge from the Wikipedia content as in [1, 14, 17, 12, 19].

DBpedia [1] is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia offers sophisticated queries against Wikipedia and to other linked datasets on the Web. The DBpedia dataset describes 1,950,000 "things", including at least 80,000 persons, 70,000 places, 35,000 music albums, 12,000 films. It contains 657,000 links to images, 1,600,000 links to relevant external web pages and 440,000 external links into other RDF datasets. Altogether, the DBpedia dataset

consists of around 103 million RDF triples. DBpedia extracts [2] RDF triples from Wikipedia informations presented in the page templates such as infoboxes and hyperlinks.

Yago [14] is a knowledge base which extends the relationships of DBpedia extending the standard RDF notation. At December 2007, Yago contained over 1.7 million entities (like persons, organizations, cities, etc.) A YAGO-query consists of multiple lines (conditions). Each line contains one entity, a relation and another entity.

DBpedia or Yago could be replaced by Wikipedia as a source of knowledge in our semantic discovery approach, although at the time of this writing these knowledge bases contain only entities (such as person and places) and not abstract concepts as the one we have in e-Learning material. In addition we don't know a priori with which properties a term can be searched, so in our domain replacing Wikipedia free-text would not be beneficial.

ISOLDE [17] is a system for deriving a domain ontology using a named-entity tagger on a corpus and combining the extracted information with Wikipedia and Wiktionary. The results show that this kind of approach works better with semi-structured information such as dictionaries.

KYLIN [19] is another project which aims to automatically complete the information presented in the Wikipedia infoboxes analyzing disambiguated text and links in Wikipedia pages.

Ponzetto et al. [12] in their work have explored information extraction on Wikipedia for creating a taxonomy containing a large amount of subsumptions, i. e. is-a relations.

**4. A Semantic Discovery Approach .** In this section we explain the process of extracting semantics from multimedia content. We tested the approach here described on e-Lecture presentations. An e-lecture is a multimedia presentation usually composed of a video with focus on the speaker, presentation slides and other textual documents which can be identified by the presenter as related source of information. Different media in a presentation are used for drawing a better picture of the content to be communicated.

Hence it is of fundamental importance to take into account the different modalities of the media. In particular, we investigate textual modality analyzing the full content of the related material such as slides or documents and the auditory modality translating it into textual which represents the most promising aspect of the data we can process. For this reason we apply automatic speech recognition (ASR) to the video soundtracks and subsequently we are interested in the STT translation (speech to text) to provide for data that can be analyzed in combination with the other textual resources such as slides, notes and other documents.

The reason of focusing on auditory and textual modality instead of visual modality is intrinsic to the nature of presentations. Unlike other domains such as movie or news, in video presentations the images in the keyframes are more or less still, usually the speaker and part of his/her presentation is captured. The scene almost never changes, the transitions being related to a change of focus (from slide to blackboard and back) or to the change of slide. So, in e-Lectures, even though low level feature recognition such as teacher gesture and facial prosody might give information about importance of certain passages, we decided not to attack this issue because we believe it would only bring us a minor added value in comparison to the knowledge we could retrieve exploring the auditory and textual modality.

Due to these considerations, we focus our research work on making better use of speech and textual content. Furthermore, relating the extracted speech and textual content with the right domain knowledge could provide another mode to tackle the semantic gap allowing more effective classification and searches on the video content.

In figure 4.1 we give a graphical illustration of the process of extracting semantic annotation from multimedia content. As input the system receives an e-Lecture presentation and based on the Wikipedia knowledge it automatically generates some descriptive labels for the multimedia content.

In the next paragraph we will describe in details how this process of automatic semantic discovery is taking place.

The explanation of our method can be split into two parts. In the first part we deal with the extraction of content from multimedia lecture materials without any regards about semantics; while in the second part we go in depth in the passages which involve discovering the semantics behind the content previously extracted from the media.

So, in order to discover the semantic present in a corpus we first have to extract and identify terms from it. Once we have the list of the words contained in each unit of the collection, we can link them through the relationships we will determine through Wikipedia. In particular the first part is about Information Extraction from the multimedia content and the second focuses on describing how through Wikipedia we can annotate the material and find semantic relationships between annotations.

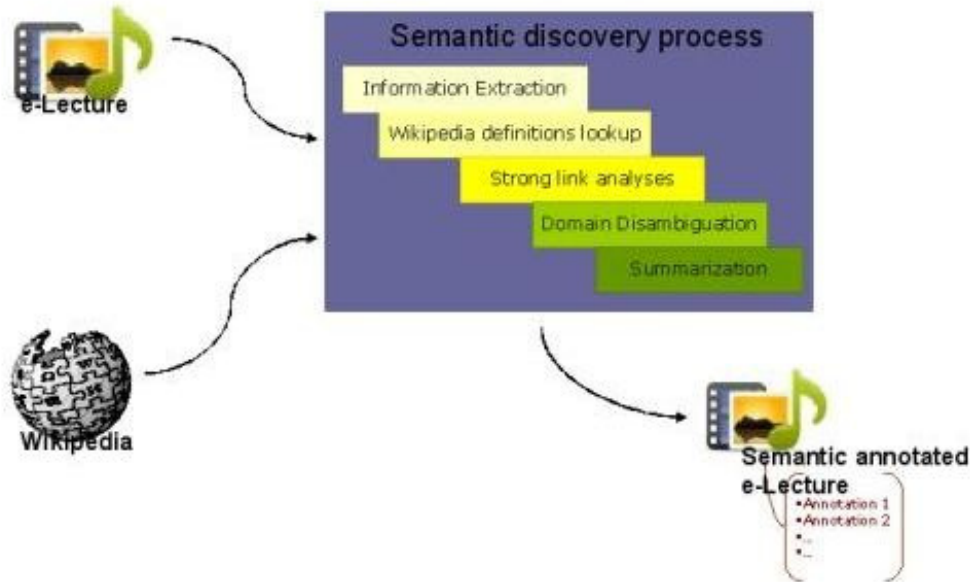


FIG. 4.1. *General Overview of the approach*

The two steps are independent in the sense that Information Extraction can be carried out in different ways while our Wikipedia module could still be used to find relations between terms. We give here an explanation of the first part only for the sake of contextualization.

**4.1. Information Extraction.** In this section we will extract and model the multimedia content through the analysis and combination of the textual and auditory modality. Using an out-of-the-box speech recognition tool we translate video speech into text, and we combine it with the text extracted from presentation slides and other textual information sources.

Secondly, we extract terms for the resulting textual resources which represent the entire content of the multimedia material. We performed Information Extraction (IE) by using Lucene (<http://lucene.apache.org/>), a state of the art tool which provides Java-based indexing and search technology using a statistical approach.

Lucene had been used in the project as a search engine for querying an unstructured e-Learning repository, but since it also provides basic APIs for analyzing text, we exploited Lucene also for extracting information from our corpus. In general term extraction tools using a statistical approach basically look for repeated sequences of lexical items.

Consequently we store the extracted information in a Lucene index that we later use for information retrieval and for extracting the most important terms out of the entire e-Lecture content. We also explored a linguistic approach based on Natural Language Processing (NLP) using other state of the art tools in the area such as GATE (<http://gate.ac.uk/>) and IBM UIMA (<http://www.research.ibm.com/UIMA/> [18]), but the approach was not suited for our use case since it is language and grammar dependent. In fact, in e-Learning the material can be in different languages, and sometimes more than one language can be combined in the same event. For instance, in some cases presentation slides are written in English while the speech is delivered in another language (e.g. in Italian). As a consequence, the work needed to adapt a linguistic approach to our needs was excessive.

Moreover, story telling does not play an important role in e-Learning—at least not in the disciplines we considered—and this makes it difficult to locate and classify atomic elements in text into predefined categories for Entity Recognition. For these reasons we chose a statistical approach and we calculated a term vector for each document in our index. The term vector contains a list of terms with their frequency in a document.

In order to calculate the term vector we had to store our multimedia material into an index. Many documents can refer to the same event. For instance, we have at least two main information sources for each event: presentation slides and video transcript.

Following our multimodal view, we modeled the entire event's multimedia material into a single document in a index. In this way the term vector calculated through Lucene is factual.

For improving the performance of this task we are currently working on the indexing phrase and in particular in the pre-processing task (e.g. cleaning the text from Italian or English stop-words and applying different language stemmers as a filter, building categorizer for improving the quality of the raking of the extracted lecture terms.).

**4.2. Semantic Extraction using Wikipedia.** In this section we explain how we enhance information retrieval based on the recognition of the most important topics and the relations between them in the content of multimedia lecture material.

Understanding what a media is about before entering in its details of the search results, which would mean watching a video presentation or reading the related material, is one of the main goal in the area of Multimedia Information Retrieval, and it could be very useful when the amount of the multimedia material grows in size. Having a way for categorizing or understanding the main concepts in the content will help in managing large multimedia repositories.

Furthermore, e-Learning users (typically students) do not have a rich understanding of the domain or of how one topic is connected to others. For this reason a tool, which has the goal of enabling access to information, should give an overview of the material content for helping end users to achieve more effective searches and acquire the needed knowledge. For example, a user looking for the term `Collection` in a Java programming class must find out about the different types of Collection such as `HashMap`, `Map` and `Set` since these terms can also be found in the lecture material, and they all mean Collection. Understanding relationships between terms in our corpus permits also us to automatically discover the important topics of an event which can be used for unsupervised classification of the material.

Our starting point for the second phase of the semantic discovery process is the list of terms which were extracted from multimedia material (video, slides, and documents) during the Information Extraction phase described in the previous section.

In Wikipedia, we look up the most important extracted terms from our corpus. The goal of this phase is to find a Wikipedia definition page for every important term and to try to extract relations to other terms by examining the hypertextual links in the page. This is done by processing links in the page. Therefore, the term of interest is found in Wikipedia, and all the links in its page are analyzed.

Pre-requirements for describing the process of extracting semantics from the Wikipedia, are the four abstract concept described below that we will be mentioned again in the second part of the section for giving a description of the process itself.

1. *Wikipedia lookup.*

For each extracted term we search for pages in Wikipedia which contain the term in their name. In Wikipedia every page is named by a string composed of topic name and topic domain. After that, we collect the links for every page. The search is made on a local copy of the English version of the Wikipedia database, but we could also reach the same result by downloading and parsing Wikipedia Web pages. We chose to maintain a copy of the database to increase the speed of the task.

2. *Strong link definition.*

We define a link to be *strong* if the page it points to has a link back to the starting page. For instance, `Rome` and `Italy` are strongly linked since the page on Rome says that it is the capital of Italy, while the page on Italy reports that Rome is the capital of the state. A minor town located in Italy will instead have a *weak* link with Italy, since in its page it will be declared that the town is in Italy but in the page for Italy the minor town will most likely not be mentioned. In our case, strong links are candidates for topics related to the searched term, and they will be used for giving user suggestions in query expansion and in the process of summary generation of Wikipedia definitions.

3. *Domain disambiguation.*

A word can have multiple Wikipedia definitions because it can assume different meanings (senses) in different domains. Among the (possibly) multiple Wikipedia definitions, we choose the one which has the most link words in common with the extracted lecture's terms. We manually checked this approach

to evaluate the accordance of the semantic expressed in the disambiguated terms with the one of the event and we find out that this is true for the majority of the cases.

#### 4. *Annotation through Wikipedia definition summarization.*

In this last step we use the extracted strong links for every important word of an event to automatically generate a summary of the word definition in Wikipedia. The summary is generated taking all the sentences from the Wikipedia definition page in which a strong link is present; usually fifty percent of the content of original definition is selected. The summary is then used for expressing the meaning of the important term. In other words, we annotate the lecture through Wikipedia terminology, and for each term we keep a brief definition.

A graphical representation of the process is given in figure 4.2.

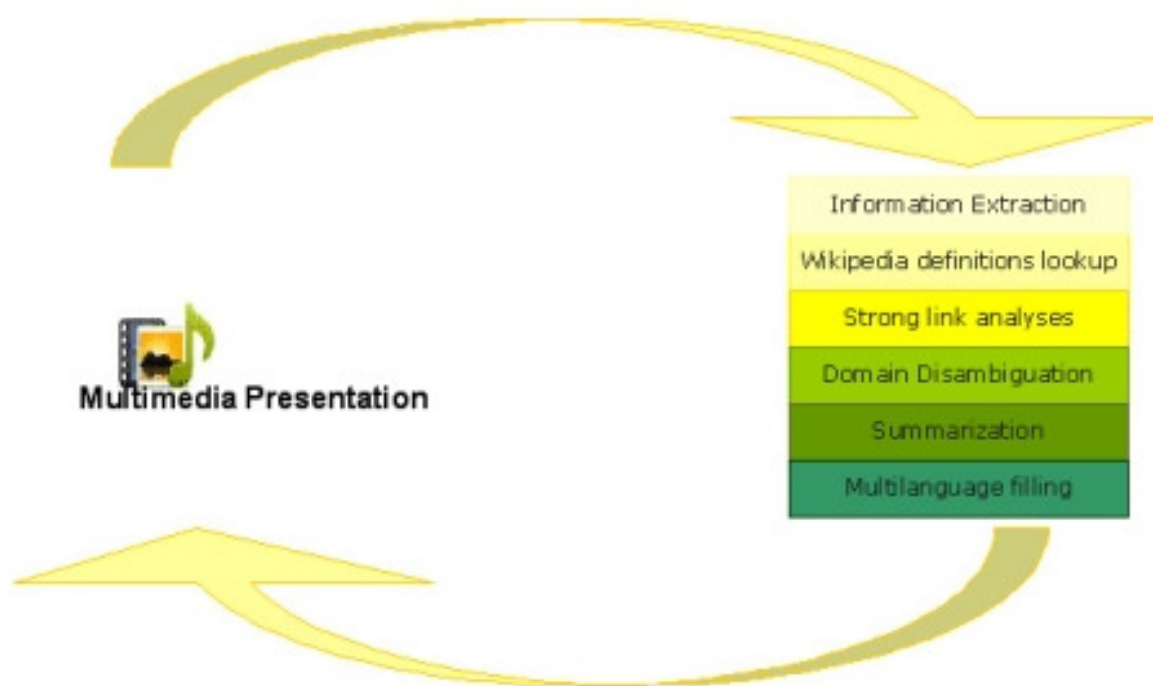


FIG. 4.2. *Semantic annotations generation process*

The figure shows that by giving e-Lectures as input and through the four steps of the semantic discovery approach we can enrich e-Lecture with semantic metadata. It follows a description of the semantic discovery process composed of four logical steps: Wikipedia definition lookup, Domain disambiguation, Strong link Analyses, and Summarization.

Suppose one of the terms extracted from the e-Lecture material is *Collection*, which is in the list of the extracted keywords. Consider a simplified list of extracted keywords (note the presence of words in more than one language!) follows: Elemento, Map, Tipo, Object, Method, Interface, computer science, Collection, Oggetto.

The first step of the algorithm is to search for all the pages which contain the term *Collection* in their name. A search in the Wikipedia database will find a relatively large number of pages that satisfy this requirement due to the different meanings the word can have. Consequently we retrieve the links of every found page. We then use the links in the disambiguation step. For instance, in the case of having a term named *Collection* in the lecture material, a Wikipedia query for the word *Collection* returns the Wikipedia disambiguation page *Collection*, which points to other pages such as *Collection(horse)*, *Collection(museum)*, *Collection(Joe Sample album)*, *Collection(agency)*, *Collection (computing)*, *Collection class*.

The second step consists in extracting the strong links from the candidate Wikipedia article for each definition. The strong links are used in the third step for calculating domain disambiguation, for support and

for automatic summarization of the content of Wikipedia entries. So in this step we extract the strong links from all the Wikipedia pages selected during the first step. For example the strong links for Collection (computing) which is one of the pages selected during the first step, are: object-oriented, class, map, tree, set, array, list. We do the same with all the pages listed in the Wikipedia disambiguation page for collection.

The third step resolves domain ambiguity. We automatically identify the right Wikipedia definition based on the domain defined by the multimedia. So we select one page among all the ones we retrieved during step one which is in concordance with the domain of interest. The disambiguation function considers for every candidate page its strong links given by step two. In particular it looks for correspondences among the link's names and the keywords extracted from the corpus. The page which has the largest number of links in correspondence with the corpus terms will be considered to be the correct one and it will be used as the disambiguated term. Supposing that Collection(computing) is the Wikipedia article is the page that responds to this requirement then every time Collection will be mention in the lecture material it will be associated with the meaning of the Wikipedia article Collection(computing). The term that has been disambiguated has the same meaning in Wikipedia and in the corpus. The result of this step is the identification of the disambiguated terms with their links. In other words this step compares the strong links of every candidate Wikipedia definition with the term vector of the lecture in exam.

During the last step we create a summary for the most important words in the lecture. Each term in the lecture has a corresponding Wikipedia definition and based on the ranking in the term vector we can identify the n most important terms in the e-Lecture. The summary creation uses the extracted strong links of the most important terms. For every one of them, we download and parse the correspondent Wikipedia. From the Wikipedia page we select only the sentences which contain a strong link and the term itself. The combination of extracted sentences permits to generate a reasonably well written article summarization.

**5. Applications.** In this section we describe some applications where our approach can be used. Many other applications are under considerations. In particular we applied the semantic discovery approach into NEEDLE [8]- Next generation sEarch engine for Digital LibrariEs -. NEEDLE is an e-learning application which aims at indexing, searching and presenting structured and unstructured multimedia data. The system provides a way to search e-learning materials through a web-based search interface.

The e-Learning materials consist of video lectures and corresponding audio tracks, PowerPoint presentation slides etc. The application's main objective is to present the structured and unstructured multimedia e-Learning materials. The users could query the NEEDLE system to search for materials of their interest.

The data i. e. video lectures and slides, for NEEDLE come from the LODÉ [5] system. LODÉ is web-based application for presenting the video lectures synchronized with presentation slides. The audio content of the video lectures from the LODÉ system is transcribed using speech recognition and speech processing tools. The text content of the transcription and PowerPoint slides are indexed and searched using NEEDLE system.

Most of the commercial search engines only offer text based search and few also provide image search. However, there is still a need for searching video, audio, graphics etc. Commercial video hosting sites like YouTube that offer search for video actually performs the search only on the meta-data (text content describing the video) attached with the video. They do not search the video/audio content. We offer textual search on audio content using the transcript in combination with the content of the documents which come with the video such as presentation slides; moreover we enhanced the search task with search suggestion based on the identification of relationship between topics in Wikipedia and we automatically extract also through Wikipedia labels for describing the most import lecture's topics.

We can summarize with the following points the features where our approach could contribute:

- *Search Suggestion and Query Expansion*

Wikipedia is used for finding topics related to the searched one. In our search user interface we show the hits for the searched string and a bunch of links to some related topics which have a correspondence in our repository. A click on one of the link will initiate a search for the occurrences of the link term in the learning material. This is done by viewing all the strong links retrieved through Wikipedia which term appear also in the event material, in this way we can suggest different search terms or topics that are connected to the first searched one.

- *Automatic Annotation*

For each occurrence displayed in the hits, we show some links to related important topics. The important topics automatically annotate the event with some terms which have a predefined meaning in Wikipedia.



In this way there are no more ambiguities in the meaning of a term used for annotation. Another advantage of the strong link identification in combination with the term vector extracted for every event is the possibility to automatically describe the most important concepts of the event.

- Automatic Summarization

The semantic discovery approach described in the previous section brought us to the individuation of the strong links for each topic. Based on them we can generate for each event annotation (topic) a brief summarization of the description of the topic in Wikipedia. A click on one of the event annotation will display the summary plus the retrieved hits for that term. In our search user interface for each event (lecture, seminar, meeting) we show the six most important words and the related summarized Wikipedia definitions.

In figure 5.1 the possible features derived by the implementation of the semantic discovery approach described above are shown. Figure 5.1 is a screen shot of NEEDLE where every lecture it is first summarized by means of a list of important topics. So the user looking at the important topics list can understand if a lecture in the hits is relevant or not for each search and in case s/he can go in depth looking at the details hits inside the lecture itself. Every hit in the lecture is composed of a brief textual description and four presentation modalities. The result can be analyzed watching part of the video where the hit have been found, listening only to the audio, only watching the associated slide or having a combine view, where video and slide are time-synchronized.



FIG. 5.1. An application of semantics extraction through Wikipedia

On the right most part of figure 5.1 a pop-up window for with a description of one of the important terms of the lecture can be viewed by clicking on the term. The description is actually the summary of the Wikipedia article which refers to the term itself. In this way there is no ambiguity with the meaning of a term used for describing the multimedia content. Below the input field designed for running new searches there is the implementation of the related topics feature. Based on the last user search, the system advises the user for related concepts. A click on one related term will initiate a new search. In this way the user can discover new connections between topics.

**6. Future directions.** One of the activities we plan for the near future is related to extending the Wikipedia module to support various languages. The multilanguage support consists in recognizing relations between terms in the corpus which are not in English. As a first step, we'll look at the links to the other instances of Wikipedia in different languages. In most cases, pages in the Wikipedia instance in one language have links to pages in many other Wikipedia instances in other languages. Since these links were created manually by the page authors, in most cases there is no ambiguity in the translation. In case a link to the target language of interest is not present, we can resort to freely available, albeit less trustable external sources for translating from and to English. The Wikipedia process described in the previous section will not change but writing language dependent processing modules such as language specific stemmers should be added to enable the comparison of the related Wikipedia content found in English with the terms contained in the multimedia content repository. Consequently we have scheduled an evaluation of the presented approach for annotating a large amount of text resources and a user based evaluation to assess if the introduction of semantic multimedia information retrieval is actually bringing an advantage to the student. We will carry out a student's performance evaluation on some topics presented in the e-Learning repository and we will compare the results with the ones we gathered last year using a text based search that was not semantically enhanced [9].

**7. Conclusion.** In this paper we described an approach to semantically annotate the content of an unstructured multimedia repository. The annotation has been done combining the terms extracted from the corpus with lexicographic relationships from Wikipedia. Wikipedia has been used as an alternative to ontologies. The content annotated in this way permits to keep track of the relations between annotations. The approach has been used for giving search suggestions in multimedia information retrieval, in multimedia annotation and for giving a brief description of the topics of the multimedia event. Our approach is domain independent, and it could in theory also be applied to different use cases where there is a need for clustering or annotation of a corpus.

## REFERENCES

- [1] K. ABERER, K.-S. CHOI, N. F. NOY, D. ALLEMANG, K.-I. LEE, L. J. B. NIXON, J. GOLBECK, P. MIKA, D. MAYNARD, R. MIZOGUCHI, G. SCHREIBER, AND P. CUDRÉ-MAUROUX, eds., *DBpedia: A Nucleus for a Web of Open Data*, vol. 4825 of Lecture Notes in Computer Science, Springer, 2007.
- [2] S. AUER AND J. LEHMANN, *What have innsbruck and leipzig in common? extracting semantics from wiki content.*, in Proceedings of the 4th European Semantic Web Conference, ESWC 2007, 2007.
- [3] M. BERTINI, A. D. BIMBO, AND C. TORNIAI, *Enhanced ontologies for video annotation and retrieval*, in MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, New York, NY, USA, 2005, ACM, pp. 89–96.
- [4] K. BONTCHEVA, D. MAYNARD, H. CUNNINGHAM, AND H. SAGGION, *Using human language technology for automatic annotation and indexing of digital library content.*, in 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL'2002), Rome, September 2002.
- [5] M. DOLZANI AND M. RONCHETTI, *Video streaming over the internet to support learning: the lode system*, in WIT Transactions on Informatics and Communication Technologies, vol. 34, 2005, pp. 61–65.
- [6] M. DOWMAN, V. TABLAN, H. CUNNINGHAM, AND B. POPOV, *Web-assisted annotation, semantic indexing and search of television and radio news*, in Proceedings of the 14th International World Wide Web Conference, Chiba, Japan, 2005.
- [7] A. EBERSBACH, M. GLASER, AND R. HEIGL, *Wiki : Web Collaboration*, Springer, November 2005.
- [8] A. FOGAROLLI, G. RICCARDI, AND M. RONCHETTI, *Searching information in a collection of video-lectures*, in Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, Vancouver, Canada, June 2007, AACE, pp. 1450–1459.
- [9] A. FOGAROLLI AND M. RONCHETTI, *Case study: Evaluation of a tool for searching inside a collection of multimodal e-lectures*, in Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, Vancouver, Canada, June 2007, AACE, pp. 3893–3900.
- [10] M. HEPP, K. SIORPAES, AND D. BACHLECHNER, *Harvesting wiki consensus: Using wikipedia entries as vocabulary for knowledge management.*, Internet Computing, 11 (2007), pp. 54–65.
- [11] L. OBRST, *Ontologies for semantically interoperable systems*, in CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management, New York, NY, USA, 2003, ACM Press, pp. 366–369.
- [12] S. PONZETTO AND M. STRUBE, *Deriving a large scale taxonomy from wikipedia*, in Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI-07), Vancouver, B.C., July 2007.
- [13] J. PORTER, *Folksonomies: A user-driven approach to organizing content*, User Interface Engineering, (2005).
- [14] F. SUCHANEK, G. KASNECI, AND G. WEIKUM, *Yago: A large ontology from wikipedia and wordnet*, Research Report MPI-I-2007-5-003, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, 2007.
- [15] J. T. TENNIS, *Social tagging and the next steps for indexing*, in 17th SIG/CR Classification Research Workshop, 2006.
- [16] M. VÖLKEL, M. KRÖTZSCH, D. VRANDEIC, H. HALLER, AND R. STUDER, *Semantic wikipedia*, in Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, May 23-26, 2006, MAY 2006.

- [17] N. WEBER AND P. BUITELAAR, *Web-based ontology learning with isolate*, in Proc. of ISWC2006 Workshop on Web Content Mining with Human Language Technologies, 2006.
- [18] C. WELTY AND J. MURDOCK, *Towards knowledge acquisition from information extraction*, The Semantic Web - ISWC 2006, (2006), pp. 709–722.
- [19] F. WU AND D. WELD, *Autonomously semantifying wikipedia*, in ACM Sixteenth Conference on Information and Knowledge Management (CIKM-07), Lisbon, Portugal, November 2007.

*Edited by:* Dominik Flejter, Tomasz Kaczmarek, Marek Kowalkiewicz

*Received:* November 23rd, 2007

*Accepted:* January 15th, 2008

*Extended version received:* June 29th, 2008