Semantic Video Classification Based on Subtitles and Domain Terminologies

Polyxeni Katsiouli, Vassileios Tsetsos, Stathes Hadjiefthymiades

Pervasive Computing Research Group, Communication Networks Laboratory,
Department of Informatics & Telecommunications, University of Athens,
Panepistimioupolis, Ilissia 15784, Greece.
{polina, b.tsetsos, shadj}@di.uoa.gr

Abstract. In this paper we explore an unsupervised approach to classify video content by analyzing the corresponding subtitles. The proposed method is based on the WordNet lexical database and the WordNet domains and applies natural language processing techniques on video subtitles. The method is divided into several steps. The first step includes subtitle text preprocessing. During the next steps, a keyword extraction method and a word sense disambiguation technique are applied. Subsequently, the WordNet domains that correspond to the correct word senses are identified. The final step assigns category labels to the video content based on the extracted domains. Experimental results with documentary videos show that the proposed method is quite effective in discovering the correct category for each video.

Key words: video classification, text classification, WordNet domains

1 Introduction

As multimedia databases gain more and more popularity, retrieving semantic information from multimedia content becomes a critical and challenging topic. In order to make efficient use of such databases it is crucial to explore efficient ways to index their content based on its features and semantics. There are many ways to perform video classification and indexing. One way is through video signal processing. Another, and in our opinion one of the most challenging approaches concerning semantic video indexing, is based on the extraction of semantics from its subtitles. Subtitles carry such information through natural language sentences.

Such approach, although may not be able to detect all video semantics (e.g., in scenes not involving spoken dialogues), can have several benefits over content classification based on visual/audio signal processing. Firstly, text and natural language processing is, in general, a more lightweight process than video and audio processing and constitutes a topic that has been studied extensively in the computational linguistics literature. Additionally, high-level semantics are more closely related to human language than to visual/audio signal features. Hence, effective text-based classification methods seem quite suitable for semantic multimedia content indexing, where applicable.

In this paper, we describe an unsupervised video classification approach, based on the WordNet lexical database and the WordNet domains, as defined in [1]. According

to our approach, a video is assigned a category label by applying natural language processing techniques on its subtitles.

The rest of the paper is organized as follows. In section 2, some related methods for video and text classification are outlined, whereas section 3 provides some information concerning the WordNet lexical database and the WordNet domains. The proposed approach is described in detail in section 4 and is followed by an experimental evaluation presented in section 5. Section 6 briefly describes the POLYSEMA project, in the context of which was performed the present work. The paper concludes with some remarks for future work and open research challenges.

2 Related Work

2.1 Video Classification

Several video classification methods based on either visual or text features have been proposed in the relevant literature.

In [2] a video indexing and summarization approach based on the information extracted from a script file in a DVD/DivX video is described. The method partitions the script in segments and represents each one as a term frequency inverse document frequency (TF-IDF) vector. The collection of these vectors is called script matrix. Two applications, video retrieval and summarization are described through the application of machine learning techniques, such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD) and clustering, to the script matrix.

The MUMIS project [3], [4] makes use of natural language processing techniques for indexing and searching multimedia content. An information extraction method based on an XML-encoded ontology is applied to textual sources of different type and in different language separately. Then, the project combines the annotations extracted from such sources into one integrated, formal description of their content.

The authors in [5] present a framework for semantic classification of educational surgery videos. Their approach consists of two phases: i) video content characterization via principal video shots, and ii) video classification through a mixture Gaussian model.

An approach for semantic video classification based on low-level features such as color, shape and motion is described in [6]. The authors adopt techniques for extracting such features from the video files and use a Support Vector Machine (SVM) classifier in order to classify them in one of the following class labels: “cartoons”, “commercials”, “cricket”, “football” and “tennis”.

Finally, in [7] a video classification method based on face and text trajectories is described. The authors use Hidden Markov Model (HMM) to classify video clips into predefined categories such as “commercial”, “news”, “sitcom” and “soap”.

2.2 Text Classification

The goal of text classification is the assignment of one or more predefined categories to a document based on its content. Text classification methods, similarly to all

classification methods, are divided into two broad categories: supervised and unsupervised.

The authors in [8] study the effectiveness of feature selection methods, such as document frequency and chi-square (χ^2), in text categorization using k-Nearest Neighbor classifier [9] and Linear Least Squares Fit mapping (LLSF) [10]. K-nearest Neighbor algorithm classifies a new object based on training samples in the feature space. It is a type of instance-based learning and can be used for regression as well. LLSF is based on a linear parametric model and uses words in the document to predict weights of categories.

The use of Support Vector Machine (SVM) in text categorization is described in [11]. SVM is a machine learning technique, used for binary classification, which performs a mapping of the input space to a feature space and constructs a hyperplane which separates the data. The author in [11] integrates dimension reduction and classification by SVM and adapts this technique to dynamic environments that require frequent additions to the document collection.

The Naïve Bayes classifier [12] is a probabilistic classifier based on the so-called Bayesian theorem and is particularly appropriate when the dimensionality of the input data is high. In text categorization the Bayes theorem is used to estimate the probability of category membership for each category and each document. Such estimates are based on the co-occurrence of categories and features in the training set. The Naïve Bayes assumes that the set of features in which the classifier is built are independent.

Decision trees [13] are an important and successful machine learning technique which can be used for classification and prediction tasks. In the structure of such a tree, the leaves represent classifications whereas the branches correspond to the combinations of attributes that leads to those classifications. In this paper, we compare the proposed method for classification with a decision tree classifier.

Finally, some hybrid indexing approaches that combine techniques from video analysis and text classification have been proposed. For instance, the authors in [14] describe a content-based image and video retrieval system with the use of embedded text. The proposed system determines the text regions of still images and video frames and applies a connected component analysis technique to them. The remaining text blocks of such analysis are used as input in an Optical Character Recognition (OCR) algorithm. The OCR output is stored in a database in the form of keywords associated with the corresponding frames.

3 WordNet and WordNet Domains

WordNet [15] is a large lexical database, not restricted to a specific domain, in which English nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms called “synsets”. Each synset contains a group of synonymous words or collocations (i.e., sequence of words that co-occur often, forming a common expression). Most synsets are connected to other synsets through a number of semantic relations which vary based on the type of the word. Specifically, noun synsets are related through *hypernymy* (generalization), *holonymy* (whole of),

hyponymy (specialization) and *meronymy* (part of) relations. Some of the relations between verb synsets are *hypernym*, *holonym*, *entailment* and *troponym*. Participial adjectives are related with verbs through the *participle of* relation. Adverbs are often derived from adjectives. Therefore, they usually contain lexical pointers to the adjectives they are derived from.

Example: **plant**	**WordNet Domains**
plant, works, industrial plant -- (buildings for carrying on industrial labor; "they built a large plant to manufacture automobiles")	**Industry**
plant, flora, plant life -- (a living organism lacking the power of locomotion)	**Biology, Plants**
plant – (something planted secretly for discovery by another; "the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant")	**Factotum**
plant -- (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)	**Theatre**

Fig. 1. Some senses of the word "plant" with their corresponding domains

The authors in [1] have created *WordNet domains* by augmenting WordNet with domain labels. A taxonomy of approximately 200 domain labels enhances WordNet synsets with additional information. Synsets have been annotated with at least one domain label, whereas a domain may include synsets of different syntactic categories as well as from different WordNet sub-hierarchies. If none of the domain labels is adequate for a specific synset, the label *Factotum* is assigned to it (almost 35% of the WordNet 2.0 synsets have been annotated with the label *Factotum*).

Fig. 1 illustrates the senses of the (noun) word "plant" and the corresponding domains for each of these senses, while a part of the WordNet domains hierarchy is depicted in Fig. 2. In the following sections we describe in detail how WordNet and WordNet domains were exploited by our video categorization approach.

4 Proposed Video Categorization Scheme

The complete approach for semantic video classification is decomposed to several steps that are described in the following paragraphs and summarized in Fig. 3.

Step 1: Text Preprocessing

During the first step of the algorithm, subtitles are segmented into sentences and a part of speech (POS) tagger is applied to the words of each phrase. Specifically, the Mark Hepple's POS tagger [16] was adopted. This step is essential in order to pick the correct meaning of each word in WordNet. Subsequently, stop words (e.g., "about", "also", "him") are removed, based on an English stop words list [17], as they carry no semantics and do not contribute to the understanding of the main text concepts.

Step 2: Keywords Extraction

In order to identify and select only the most important and relevant subtitle words for further classifying the video, we implemented the TextRank [18] algorithm. This is a well known algorithm among the text classification community with proven performance. Specifically, TextRank is a completely unsupervised graph-based ranking model, used for text applications such as keywords extraction and text summarization. The TextRank algorithm builds a graph that represents the text and applies to the graph vertices a ranking algorithm derived from Google's PageRank algorithm [19]. Then, the vertices are sorted in reverse order of their score and the top *T* vertices are extracted for further processing. The number of keywords extracted is based on the size of the text. Specifically, *T* is set to a third of the number of vertices in the graph.

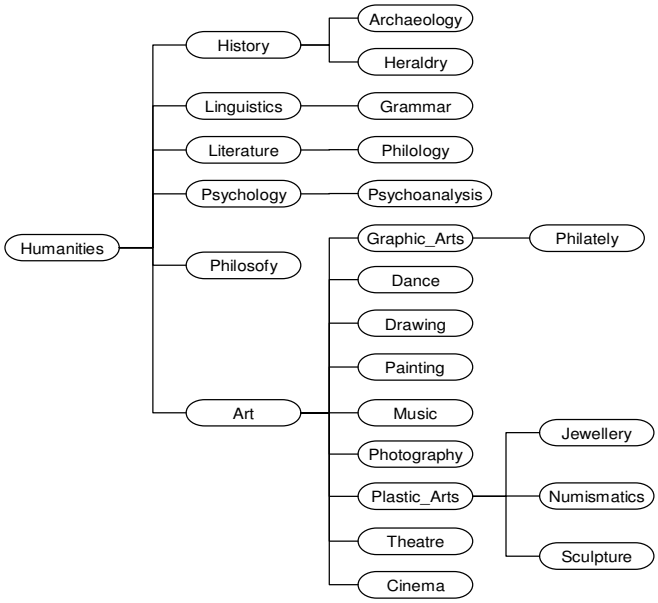

Fig. 2. Extract of WordNet Domains hierarchy

Step 3: Word Sense Disambiguation

In order to improve the effectiveness of our approach, we applied a Word Sense Disambiguation (WSD) method. Most words in natural language are characterized by polysemy, i.e., they have many possible meanings, called senses. WSD is the task of finding the correct sense of a word in a specific context. To assign a sense to each word in the text we used the WSD algorithm presented in [20]. This algorithm is an adaptation of Lesk's algorithm [21] for WSD. According to Lesk's algorithm, which is based on glosses found in traditional dictionaries, a word is assigned the sense whose gloss shares the largest number of words with the glosses of the other words in the context of the word being disambiguated. The authors in [20] extend Lesk's algorithm using WordNet to include the glosses of the words that are related to the

word being disambiguated through semantic relations, such as hyponym, hypernym, holonym, troponym and attribute of each word. Suppose, for example, that we want to disambiguate the word '*bank*' in the phrase '*he sat on the bank of the river*'. While Lesk's algorithm compares the glosses of the word '*bank*' with those of '*river*' and '*sat*', the authors in [20] compare the glosses of the senses of the word '*bank*' with the glosses of the hyponyms, hyponyms or holonyms of the other surrounding words.

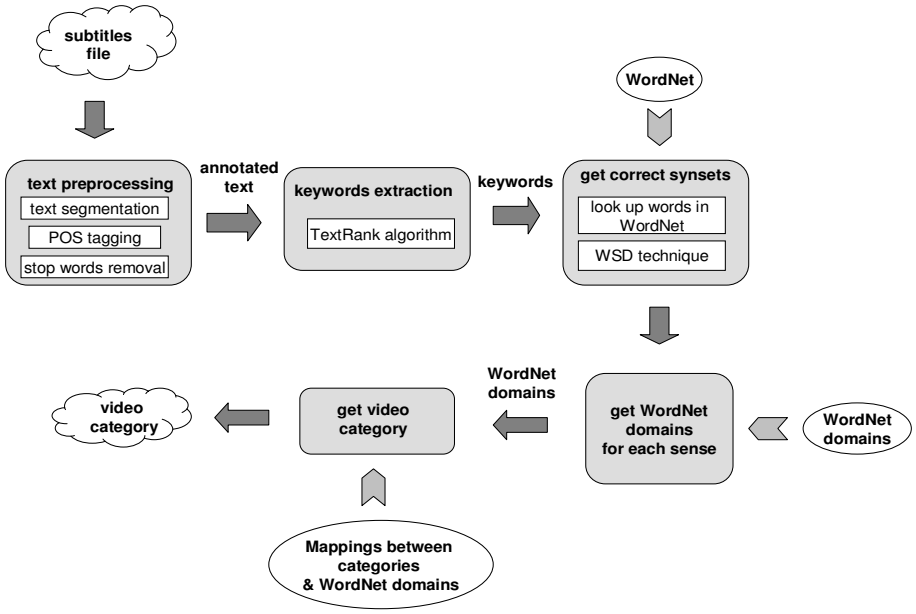

Fig. 3. Overall approach for video classification

Step 4: WordNet Domains Extraction

Having identified the correct synsets for each of the keywords extracted in Step 2 we make use of the WordNet domains to derive the domains which these synsets correspond to. Subsequently, we calculate the occurrence score of each domain label (i.e., how many times the label appears in the text) and sort them in decreasing order. We extract the WordNet domains with the highest occurrence score, as these domains will affect the category assignment in the last step.

Step 5: Definition of correspondences between category labels and WordNet domains

In order to choose the most appropriate class label for each video, we adopted the following procedure, which defines mappings between WordNet domains and category labels (see Fig. 4). First, we looked up in WordNet the senses related to each category label. Then, we obtained the WordNet domains that correspond to the senses of each category. Subsequently, we calculated for each category the occurrence score of each of the derived domains and sorted them in decreasing occurrence order.

Input:

C: all category labels

WND: all mappings between WordNet synsets and WordNet domains

Output:

D: A hash table with hash function f: V→ L, where L is a list of WordNet domains that correspond to each category label in decreasing occurrence order

For each category $c \in$ C **do**

$D_c \leftarrow \varnothing$

Let *S* be the set of senses of category c as well as the senses related with c through hypernym & hyponym relations

For each sense s $\in S$ **do**

Let *W* be the WordNet domains that correspond to sense s (obtained from WND)

$D_c \leftarrow D_c \cup W$

End for

Calculate the occurrence score of each domain in D_c

Let D_c' be the list which contains the elements of D_c in decreasing occurrence order

D[c] $\leftarrow D_c'$

End for

Fig. 4. Algorithm for the definition of correspondences between WordNet domains and category labels

Step 6: Category label assignment

The final step of the proposed approach deals with assigning a category label to the video entity. During this step the top-ranked WordNet domains produced by the process of Fig. 4, are compared to the set of the WordNet domains extracted in Step 4 (see Fig. 5). Specifically, let *C* be the set with all the category labels and *D* the set of all the WordNet domains that correspond to each category label, as produced by the algorithm of Fig. 4 (1).

$$D = \bigcup_{c \in C} \{D_c'\} \tag{1}$$

Suppose that W_v is the ordered list of the WordNet domains for a video *v,* as extracted during Step 4 of the proposed method. The process continues by checking which category $c \in C$ satisfies equation (2), and classifies video *v* under the category *c*.

$$D_c'[0] = W_v[0] \tag{2}$$

Input:
D: A hash table with hash function f_D: V→ L, where L is a list of WordNet domains that correspond to each category label in decreasing occurrence order
V: A set of the videos being classified
C: A set of all category labels
Wv: The ordered list of the WordNet domains for video $v \in V$
Output:
A: A hash table with hash function f_A: V → C
For each video $v \in V$ **do**
  Let Cv ← ∅ be the set of all candidate category labels for video v
  j ← 0
  i ← 0
  Repeat
    For each category $c \in$ C **do**
      Dc ← D[c] /*Dc contains all the WordNet domains for category c in decreasing occurrence order*/
      If Dc[i] = Wv[j] **then**
        A[v] ← c
        Cv ← Cv ∪ {c}
      End if
    End for
    If |Cv| = 1 **then break** /*assign category c to video v*/
    If |Cv| > 1 **then**
      i ← i + 1
      j ← j + 1
      C ← Cv
    End if
    If Cv = ∅ **then** j ← j +1
  Until A[v] ≠ null
End for

Fig. 5. Category label assignment algorithm

In case there are more that one such categories (e.g., c_i and c_j), the method compares the second elements of the corresponding sets, and so on. If, on the other hand, there is no category $c \in C$ that satisfies (2), we continue by checking which of the category labels satisfies the following equation:

$$D_c'[0] = W_v[1] \tag{3}$$

The method continues as described above until a category label in assigned to the video.

5 Experimental Evaluation

In order to assess the effectiveness of the proposed approach we used subtitles of documentaries[1]. In general, it is considered easier to classify documentaries since they are usually restricted to a specific domain and usually contain narrative. Table 1 presents some indicative statistical information concerning the subtitle files used in the experiments. The last column of the Table 1 indicates the number of domains which were extracted during the Step 4 of the proposed method and have occurrence score greater than 1. Moreover, approximately 44% of all the WordNet domains extracted from each video are assigned the label 'Factotum'.

Table 1. Indicative statistical information about subtitles

Subtitles file name	Video duration (min:sec)	# of words	# of non stop words	# of keywords	# of domains
THRAKI	36:08	3430	1696	319	63
DISASTER KRUSK	47:04	5583	2577	419	61
OPUSDEIEN	47:58	6747	2992	469	55
KARAMANLIS	50:55	4202	1914	373	46
ERESSOS	28:17	2204	1081	213	46
DIONPELL	39:54	3156	1544	262	57
THE PRICE OF WAR	47:23	5139	2266	409	59
CACOYIANNIS	51:25	5388	2174	364	51
CANNIBALISM	46:21	5484	2502	442	59
DUELEN	22:42	3090	1255	236	34

In this paper we have focused on the most popular TV broadcast types (a.k.a. genres) for documentaries, namely Geography, History, Animals, Politics, Religion, Sports, Music, Accidents, Art, Science, Transportation, Technology, People and War.

Three human subjects watched separately the documentaries and classified them under these categories. We chose as expert assignments for each video the opinion of the majority, though there was no important disagreement between their decisions.

We based the evaluation on *Classification Accuracy (CA)*, a commonly used quality metric in Information Retrieval, which reflects the proportion of the classifier's correct category assignments that agree with the user's assignments (we use the term Classification Accuracy instead of Precision, because all videos are classified). Moreover, we used the *Recall* and *F-measure* performance measures to evaluate the classification results for each individual category. Recall reflects the fraction of the correct category label assignments for each category among all the expert assignements of this category, whereas F-measure represents the harmonic mean of Classification Accuracy and Recall.

During the first step of the evaluation process we calculated the mappings between the WordNet domains and the aforementioned category labels by applying the

[1] The documentaries were provided by the Lumiere Cosmos Communications company, http://www.lumiere.gr

algorithm presented in Fig. 4. Table 2 shows the WordNet domains which characterize each category (only the top-ranked WordNet domains are shown).

Table 2. Correspondences between WordNet domains and category labels

Category	**Top rank WordNet Domains**
Geography	geography
Animals	animals, biology, entomology
Politics	politics, psychology
History	history, time_period
Religion	religion
Transportation	transport, commerce, enterprise
Accidents	transport, nautical
Sports	sport, play, swimming
War	military, history
Science	medicine, biology, mathematics
Music	music, linguistics, literature
Art	art, painting, graphic_arts
Technology	engineering, industry, computer_science
People	sociology, person

We applied the proposed method to the subtitles of 36 documentaries and we calculated the CA value for all categories. Furthermore, we calculated the values of CA, Recall and F-measure for each category separately. The results were compared to those obtained from a classifier of the WEKA tool [22], [23]. WEKA is an open source software with a large repository of machine learning algorithms for data mining tasks including classification, clustering and attribute selection.

We chose the decision tree classifier J4.8, found in the WEKA repository, which is WEKA's implementation of the decision tree learner C4.5. We used the training set as the evaluation method for J4.8. This indicates that the results obtained from the training data are optimistic in comparison with what might be obtained using cross - validation [22]. Before we evaluate the performance of the classifier we removed the stop words. Table 3 presents the total CA values of the proposed method and J4.8 classifier (i.e., computed over all subtitles), whereas Table 4 presents the values of CA, Recall and F-measure for each classifier and for each category separately.

Table 3. Performance comparison of classifiers

	Classification Accuracy
Proposed method	69,4%
J4.8	89,18%
Proposed method (rank correlation coefficient)	58,3%

The J4.8 classifier results show how well the derived model performs on the training set. Comparing the results of our proposed method with J4.8 classifier, it is clear that the results obtained are very promising since it achieved an accuracy value of 69.4%. The distance between the CAs of J4.8 and our approach was somewhat expected since our method performs unsupervised classification. In the future we plan

to compare the performance of our approach with that of other unsupervised methods classifications.

Table 4. Evaluation metrics of each classifier for each category separately

	Proposed Method			J4.8			Proposed Method (rank correlation)		
	CA	Recall	F-measure	CA	Recall	F-measure	CA	Recall	F-measure
Animals	1	1	0.857	0.75	1	1	1	1	1
Geography	0.5	1	1	1	1	0.667	0.36	0.8	0.5
Politics	0.75	1	1	1	1	0.857	1	0.667	0.8
History	0.5	0.125	0.947	0.9	1	0.2	-	-	-
Accidents	1	0.5	1	1	1	0.667	0.5	0.5	0.5
People	0.6	0.75	0.667	1	0.5	0.667	0.4	0.5	0.44
War	0.667	1	0.889	0.8	1	0.8	0.8	1	0.889
Religion	1	1	1	1	1	1	1	1	1
Music	1	1	0.667	0.5	1	1	1	1	1
Art	0	0	0	0	0	0	-	-	-
Transport	1	1	0	0	0	1	-	-	-

In order to further assess the performance of the category label assignment algorithm (Fig. 5) we compared it to a rank correlation coefficient. Specifically, we used an extension of the Spearman's footrule distance that computes the correlation of top-k lists (since the lists *Wv* and *Dc'* are not of equal length and they are not permutations of the same set of values). The exact algorithm for the top-k list can be found in [24]. The coefficient used is given by equation (4), divided by the maximum distance value, $k \cdot (k+1)$, in order to normalize the values. In this equation, $\tau_1 = Wv$ and $\tau_2 = Dc'$. The parameter l is assigned the value $k+1$ as advised by the authors in [24].

$$F^{(l)}(\tau_1, \tau_2) = \sum_{i \in D_{\tau 1} \cup D_{\tau 2}} |\tau_1'(i) - \tau_2'(i)| \quad (4)$$

The result was that none of the videos was assigned one of the category labels History, Art and Transport. Moreover, 8 out of the 36 label assignments were modified with respect the results of our algorithm, with none of the modified values agreeing with the users' assignment. On the other hand, 4 out of these 8 videos were classified correctly by our algorithm but were misclassified by the Spearman's footrule coefficient. Hence, the classification accuracy of the new algorithm decreased to 58.3%.

6 A Note on the POLYSEMA Platform

The work described in this paper, has been carried out in the context of the POLYSEMA project. This project develops an end-to-end platform for interactive TV services, including a novel residential gateway architecture that is capable of

providing intelligent iTV services by exploiting the metadata of the broadcast transmission. The focus of the project, regarding semantics management, is threefold:

1. Development of semantics extraction techniques for automatic annotation of audiovisual content. These techniques are mainly applied to subtitles and involve natural language processing techniques. Three kinds of techniques are currently investigated: video summarization, domain ontology learning and video classification.
2. Development of a personalization framework for iTV services, implemented with Semantic Web technologies. A core part of this framework is an MPEG-7 ontology, which specializes and modifies the one presented in [25].
3. Development of a tool with a graphical user interface for the manual annotation of video and the creation of MPEG-7 metadata (in XML and OWL format).

The present work is part of the first activity in this list. More details on the project objectives and technical approach followed can be found in [26] and [27].

7 Conclusions - Future Work

The recent explosion in the amount of available multimedia (and especially audiovisual) data repositories has increased the need for semantic video indexing techniques. In this paper, we discussed an innovative method for unsupervised classification of video content by applying natural language processing techniques on their subtitles. The experimental results using documentaries indicate that the proposed method is very promising, especially given the fact that no training phase is required.

However, there are many improvement and extension points we are currently working on. Among them is the application of the method on a per video segment basis. Specifically, each video stream can be divided into video segments (similar to chapters in DVD movies) and the subtitles of each segment can be processed with the support of domain ontologies. The ultimate goal of such process is the classification of each segment based on its semantic content. Additionally, we are currently comparing the performance of our approach to other text classification algorithms (mainly unsupervised approaches). Moreover, in order to improve the effectiveness of the proposed method for movies (which as already mentioned, usually involve many domains), it is essential to define some knowledge domains (i.e., additional WordNet domains) more close to the movie classification (e.g., violence terms). This task can be performed either manually with the aid of linguistics researchers, or with automatic term clustering methods. Classification of movies is expected to be more challenging, since the subtitles represent dialogues and not monologues, that are typically found in documentaries. Another research direction is on substituting the keywords extraction algorithm (i.e., TextRank) with some feature selection techniques used in data mining applications (see also [8]).

The potential and the added value of the proposed approach is further increased if one combines it with existing speech to text engines and algorithms, which can be used for creating subtitles, whenever they are no available. Some of the challenges that need to be addressed in such process include the detection of speech disfluencies

and the removal of invalid words (e.g., slang language) and phonemes frequently found in human speech.

Acknowledgements

This work was partially funded by the Greek General Secretariat for Research and Technology (GSRT) and the EU, under the Operational Program "Information Society".
The authors would also like to thank the anonymous reviewers for their valuable comments.

References

1. Bentivogli, L., Forner, P., Magnini, B., Pianta, E.: Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In Proceedings of COLING Workshop on Multilingual Linguistic Resources, Geneva Switzerland (2004) 101-108
2. Yi, H.R., Deepu, R., Chia, L.T.: Semantic Video Indexing and Summarization Using Subtitles. Lecture Notes on Computer Science, Vol. 3331. Springer-Verlag Berlin Heidelberg New York (2004) 634-641
3. Declerck, T., Kuper, J., Saggion, H., Samiotou, A., Wittenburg, P., Contreras, J.: Contribution of NLP to the Content Indexing of Multimedia Documents. Lecture Notes on Computer Science, Vol. 3115. Springer –Verlag Berlin Heidelberg New York (2004) 610-618
4. Reidsma, D., Kuper, J., Declerck, T., Saggion, H., Cunningham, H.: Cross document annotation for multimedia retrieval. In EACL Workshop Language Technology and the Semantic Web (NLPXML) Budapest (2003)
5. Hangzai Luo, Jianping Fan, Jing Xiao, Xingquan Zhu, Semantic principal video shot classification via mixture Gaussian. In Proc. of IEEE International Conference on Multimedia & Expo. Vol.1, Baltimore, MD. (2003)
6. Suresh, V., Mohan, K.C., Swamy, K.R., Yegnanarayana, B.: Content-based Video Classification Using Support Vector Machines. In ICONIP-04, Calcutta, India (2004) 726-731
7. Dimitrova, N., Agnihotri, L., Wei, G.: Video classification based on HMM using text and faces. In European Signal Processing Conference. Tampere Finland (2000)
8. Y. Yang, and J. O. Pedersen. A Comparative Study on Feature Selection in Text Categorization. In Proceedings of the Fourteenth International Conference on Machine Learning.. D. H. Fisher, Ed. Morgan Kaufmann Publishers, San Francisco, CA (1997) 412-420
9. Yang, Y.: Expert network: Effective and Efficient learning from human decisions in text categorization and retrieval. In 17th Ann Int. ACM SIGIR Conference on Research and Development in Information Retrieval (1994) 13-22
10. Yang, Y., Chute, C.G.: An example-based mapping method for text categorization and retrieval. ACM Transaction on Information Systems (TOIS) (1994) 253-277
11. Kwok, J. T.-Y.: Automated Text Categorization Using Support Vector Machine. In Proceedings of the International Conference on Neural Information Processing. Kitakyushu Japan (1998) 347-351
12. Mitchell, T.: Machine Learning, McGraw Hill (1996)

13. Lewis, D., Ringuette, M.: A comparison of two learning algorithms for text categorization. In Symposium on Document Analysis and Information Retrieval. University of Nevada Las Vegas (1994)
14. Misra, C., Sural, S.: Content Based Image and Video Retrieval Using Embedded Text. Lecture Notes on Computer Science. Vol. 3852. Springer-Verlag Berlin Heidelberg New York (2006) 111-120
15. Fellbaum, C. (ed.): Wordnet: An Electronic Lexical Database. Language, Speech and Communication. MIT Press (1998)
16. Hepple, M.: Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based Part-of-Speech Taggers. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Hong Kong (2000)
17. Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Reading, Pennsylvania (1989)
18. Mihalcea, R., Tarau, P.: TextRank: Bringing Order into Texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Barcelona Spain (2004)
19. Brin S., Page L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems. Vol 30. No. 1-7 (1998) 107-117
20. Banerjee, S., Pedersen, T.: An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In the Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-02) Mexico City, Mexico (2002)
21. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from a ice cream cone. Proceedings of the 5th Annual International Conference on Systems Documentation (1986)
22. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques. 2nd Edition, Morgan Kaufmann, San Francisco (2005)
23. Weka 3 – Data Mining with Open Source Machine Learning Software in Java. http://www.cs.waikato.ac.nz/ml/weka/
24. Fagin, R., Kumar, R., Sivakumar, D.: Comparing Top k Lists. In Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (2003)
25. Tsinaraki, C., Polydoros, P., Christodoulakis, S.: Interoperability Support for Ontology-Based Video Retrieval Applications. Lecture Notes on Computer Science, Vol. 3115. Springer-Verlag Berlin Heidelberg New York (2004) 582-591
26. Papadimitriou, A., Anagnostopoulos, C., Tsetsos, V., Paskalis, S., Hadjiefthymiades, S.: A Semantics-aware Platform for Interactive TV Services. In the Proceedings of the 1st International Conference on New Media Technology (I-MEDIA '07) Graz Austria (2007)
27. The POLYSEMA Project: Multimedia Applications Supported by Semantics, URL: http://polysema.di.uoa.gr