Relevance of ASR for the Automatic Generation of Keywords Suggestions for TV programs

Véronique Malaisé[1] Luit Gazendam[2] Willemijn Heeren[3] Roeland Ordelman[3,4] Hennie Brugman[5]

(1) VU University Amsterdam, (2) Telematica Instituut, Enschede

(3) University of Twente, Enschede

(4) Netherlands Institute for Sound and Vision, Hilversum

(5) MPI for Psycholinguistics, Nijmegen

vmalaise@few.vu.nl

Résumé. L'accès aux documents multimédia, dans une archive audiovisuelle, dépend en grande partie de la quantité et de la qualité des métadonnées attachées aux documents, notamment la description de leur contenu. Cependant, l'annotation manuelle des collections est astreignante pour le personnel. De nombreuses archives évoluent vers des méthodes d'annotation (semi-)automatiques pour la création et/ou l'amélioration des métadonnées. Le project CATCH-CHOICE, fondé par NWO, s'est penché sur l'extraction de mots clés à partir de resources textuelles liées aux programmes TV destinés à être archivés (péritextes), en collaboration avec les archives audiovisuelles néerlandaises, Sound and Vision. Cet article se penche sur la question de l'adéquation des transcriptions de Reconnaissance Automatique de la Parole développés dans le projet CATCH-CHoral pour la génération automatique de mots-clés : les mots-clés extraits de ces ressources sont évalués par rapport à des annotations manuelles et par rapport à des mots-clés générés à partir de péritextes décrivant les programmes télévisuels.

Abstract. Semantic access to multimedia content in audiovisual archives is to a large extent dependent on quantity and quality of the metadata, and particularly the content descriptions that are attached to the individual items. However, the manual annotation of collections puts heavy demands on resources. A large number of archives are introducing (semi) automatic annotation techniques for generating and/or enhancing metadata. The NWO funded CATCH-CHOICE project has investigated the extraction of keywords from textual resources related to TV programs to be archived (context documents), in collaboration with the Dutch audiovisual archives, Sound and Vision. This paper investigates the suitability of Automatic Speech Recognition transcripts produced in the CATCH-CHoral project for generating such keywords, which we evaluate against manual annotations of the documents, and against keywords automatically generated from context documents describing the TV programs' content.

Mots-clés : Extraction de mots clés, Reconnaissance Automatique de la Parole, Documents Audiovisuels.

Keywords: Keyword extraction, Automatic Speech Recognition, Audiovisual Documents.

1 Introduction

Improving semantic access to multimedia content in audiovisual (AV) archives is to a large extent dependent on quantity and quality of the metadata, and particularly the content descriptions that are attached to the individual items. However, given the growing amount of materials that are being created on a daily basis and the digitization of existing analogue collections, the traditional manual annotation of collections puts heavy demands on resources, especially for large audiovisual archives. One way to address this challenge, is to introduce (semi) automatic annotation techniques for generating and/or enhancing metadata : either by doing semantic analysis on the AV items themselves (e.g., automatic speech recognition technology (Kohler *et al.*, 2008) or visual concept detection (Smeulders *et al.*, 2000)) or on textual resources related to the AV items : context documents (e.g., online TV guides, broadcaster's websites or collateral data such as subtitles (Lespinasse & Bachimont, 2001)).

This paper investigates the suitability of Automatic Speech Transcripts as a source forautomatically generating keywords suggestions. We evaluated the results extracted from these documents (1) against manual annotation of the same set of data, and (2) against keywords suggestions extracted from more canonical written documents : online Websites describing the semantic content of TV programs form our corpus. This investigation combines research from two NWO-CATCH projects : CHOICE [1], and CHoral [2], in collaboration with the Netherlands Institute for Sound and Vision, the Dutch NAtional Audiovisual Archives [3].

In addition to large quantities of audiovisual content (a.o. from Dutch public broadcasters) that are flowing in digitally on a daily basis, Sound and Vision is retrospectively digitizing thousands of hours of archival content. The automation of semantic annotation seems necessary to guarantee access to the content. In order to investigate how information from broadcasters' websites can be used for (semi) automatic content annotation, Sound and Vision started archiving these websites in the LiWa project [4]. In CHOICE, the textual content from such broadcasters' websites was used for recommending cataloguers terms from the Sound and Vision thesaurus (GTAA, see (Gazendam *et al.*, 2006), (Brugman *et al.*, 2008)).

Another source for keyword recommendation is the automatic analysis of the content itself, such as speech transcripts. Indeed, CHoral focuses on the use of automatic speech recognition (ASR) technology for indexing audiovisual documents, generally referred to as Spoken Document Retrieval (SDR). Speech recognition produces a word-level, textual representation of the audio, which –after conversion to an index– can be used for word-level search that retrieves the exact fragment in which a match for the query was found. In addition to using ASR to generate a time-stamped index for spoken word documents, it may also be useful for other purposes, such as keyword recommendation.

Whereas websites show a high level of abstraction over the content of the program (a few lines to some pages of textual description, underlining the broadcaster's opinion about the main focus of the program), the ASR transcripts are long documents, which do not follow the strict rules of written language, as they transcribe *speech*, and they are close to the *content level* of

1. `http://www.nwo.nl/CATCH/CHOICE`, funded by the Netherlands Organization for Scientific Research (NWO).
2. `http://www.nwo.nl/CATCH/choral/`, funded by the Netherlands Organization for Scientific Research (NWO) (2006-2010)
3. `http://www.beeldengeluid.nl`
4. `http://www.liwa-project.eu/`

the program itself. Indexes for audiovisual collections generated on the basis of ASR output are already being used as an alternative to standard metadata for searching spoken word documents.

In the use-case of a disclosed set of audiovisual documents provided by Sound and Vision, we discuss in this paper the usefulness of ASR transcripts for automatically generating keyword-based indexes for audiovisual documents. We use two sources of evaluation : (i) evaluation of the ASR-based keywords in terms of precision and recall against the manual annotations ; (ii) a comparison of this set with keywords suggestions generated from websites related to the TV programs (the approach investigated by CHOICE so far). This enables us to check the compatibility and complementarity of the two approaches for automatically generating multimedia documents' meta-data.

The remainder of the paper is organized as follows : in section 2 we present our pipeline for extracting keywords from textual resources related to TV programs. We then detail the ASR method that was used for this experiment in section 3. The experiment is presented in section 4 : the proposition of keywords that correspond to the in-house thesaurus of the Dutch AV archives, based on strings matched in the ASR transcript. We conclude the paper in section 5. We conclude the paper in section 5.

2 Keywords Recommendation

The extraction of keywords from textual resources has been implemented in several tools and platforms, which belong roughly to three categories : fully manual, semi-automatic and fully automatic. From a documentalist's point of view it is crucial that the extraction process does not put extra demands on the (already heavy) work load. Hence, keyword extraction systems that rely on fully manual input, such as the Annotea (Kahan & Koivunen, 2001) system, are not considered for our use case. Instead, the aim is to provide documentalists with a pre-processed list of recommended keywords from a controlled vocabulary. Semi-automatic annotation pipelines do not seem to be suitable either, although they keep a human intervention in the loop, as we wish to do. Tools from this category (like Amilcare (Ciravegna & Wilks, 2003)) help annotating the *textual documents* faster and better, but it would take an additional effort for Sound and Vision's cataloguers to annotate texts they will not archive. In our context, the documents to be annotated are the TV programs the texts *refer to*, not the texts themselves.

We opted for a fully automatic approach, such as used in the KIM system (Kiryakov *et al.*, 2005), that generates a list of keyword suggestions without human intervention. The cataloguers only need to *select* the items they consider relevant in the list. The top level of KIM's annotation ontology however, is fixed and cannot be changed. Sound and Vision's thesaurus, the GTAA, which we want our annotations to refer to (see section 4), has its own specific structure that does not comply to the model used in KIM. We therefore implemented the keyword extraction tool in GATE and co-designed the Apolda plugin [5], described in the next section.

2.1 CHOICE's Keyword Extraction Setup

CHOICE's annotation pipeline, or automatic keywords extraction pipeline, consists of two main parts : a lookup module that generates annotations referring to strings in the processed text and a

5. http://apolda.sourceforge.net/

ranking module. The lookup, done with the Apolda plug-in mentioned above, is based on simple string matching linked with some heuristics, like preferring the longest possible match for an annotation. Because the lookup does not include linguistic rules, it is language and domain-independent, and can also apply to textual resources that are not conform to standard written language (blogs on Internet or ASR transcripts for example). However, as simple string matching leads to synonymy and polysemy problems, ranking algorithms have to be implemented to filter out errors in the keywords proposition list. We have shown in (Malaisé *et al.*, 2007) that a ranking algorithm can be used as a Word Sense Disambiguation module. After comparing different methods based on the thesaurus' graph structure (Gazendam *et al.*, 2009), we found out that a classic tf.idf ranking performed equally good (although the generated lists are different). Therefore, for this experiment, we use the classic tf.idf ranking. tf measures the frequency of a possible keyword in a document, df is its frequency in the corpus taken into account (in our case a disclosed subset of the audiovisual archives, described in section 4).

3 Automatic Speech Recognition Setup

An alternative to broadcasters' websites for generating keyword recommendation is to use speech transcripts of the audiovisual content. Speech in audiovisual documents may represent the content of an item very well and, once converted from audio samples into a textual representation, could serve as a useful source for keyword selection. Different types of speech transcripts can be thought of : they may be generated manually in the production stage of the AV document (e.g., teleprompts, scripts, summaries) or afterwards (subtitles), or generated automatically using automatic speech recognition technology. In our use case, the Dutch SHoUT (Huijbregts, 2008) speech recognition toolkit is deployed. Its performance on the TRECVID2007 [6] data set that is used for our keyword selection experiments, is elaborately discussed in (Huijbregts *et al.*, 2007), but we will give a summary here.

The SHoUT transcription setup consists of Speech Activity Detection, speaker segmentation and clustering, and multi-pass speech recognition. During Speech Activity Detection speech is separated from non-speech segments. Especially in the broadcast domain, this step is important for filtering out e.g., music, environmental sounds, and long stretches of silence, before the audio is processed. Within the segments identified as speech, speaker segmentation and clustering is used to group intervals of speech spoken by the same speaker. Moreover, those clusters can be employed for model adaptation for, e.g., male versus female speakers, or individual speakers. The automatic speech recognition generates a 1-best transcript, in this case using acoustic models trained on 79 hours of read and spontaneous speech, and language models trained on 500 million words, mainly from newspaper texts. For each individual item, topic specific language models were created by interpolating item-specific models with a general background model. With state-of-the-art systems for speech activity detection and speaker clustering the automatic speech recognition output showed an average word error rate (WER) of 60.6%. Note that in comparison with ASR performance of WERs between 10 and 20% on broadcast news, error rates on the heterogeneous speech in the TRECVID data set (historical content, background noise, spontaneous speech) are much higher. This is taken to be due to remaining mismatches between the data and the models that were used.

Improvement of the ASR performance in such heterogeneous domains is not trivial, however,

6. The TRECVID 2007 data consists of content from the (Dutch) Sound and Vision Academia collection.

because it would require large amounts of text data that both match the characteristics of spontaneous speech and the topic of the broadcasts, which is hard to find. System performance is comparable to ASR performance on other (non-Dutch) collections of more spontaneous speech materials, such as the National Gallery of the Spoken Word collection (Hansen *et al.*, 2005), and the MALACH interview collection (Byrne *et al.*, 2004).

4 Experiment

To evaluate the suitability of ASR transcripts for annotating TV programs with keywords from a given thesaurus, we ran the experiment detailed below.

4.1 Material

The Academia collection of TV programs [7] has been cleared from intellectual property rights by Sound and Vision in order to create an open accessible collection for educational and research purposes. This collection is a subset of the Sound and Vision's catalogue, and has therefore been manually annotated by professional cataloguers. This set has been subject to the TREC-VID 2007 competition. We use in this experiment a set of 110 documents from this Academia collection, which has been processed by the ASR system described in section 3. In a second step, to compare the results of the keywords suggestion based on this source with our previous experiments, we also looked for the corresponding set of Website descriptions. Unfortunately, we could only find 13 Website texts corresponding to Academia's TV programs as some broadcaster's websites do not archive the descriptions of their programs. This low number is due to the variety of genres that can be found in this corpus ; our previous experiments were taking only documentaries into account, which have extensive textual descriptions that can date back several years. Therefore the suitability of the websites' text as source for generating keywords is highly correlated to the TV program's genre.

The keywords lists that are generated correspond to the controlled vocabulary used for manual indexing at Sound and Vision : the GTAA thesaurus. GTAA is a Dutch acronym for "Common thesaurus for Audiovisual Archives" ; it contains about 160 000 terms, divided in different facets. In this experiment, we only take into account keywords from the facet describing the subject of a TV program, which contains about 3800 preferred terms and 2000 non-prefered terms, *i.e.* other words that correspond to the same notion or term. We added to these sets a list of synonyms computed from online dictionaries and the singular forms of the terms (which are mostly represented in the plural form, following the recommendation of the ISO 2788 :1986), from the CELEX lexical resource (Baayen *et al.*, 1995). Our documents and thesaurus are in Dutch.

4.2 Evaluation Metrics

The evaluation of the keywords suggestion based on the strict (i.e. classic) measure of precision and recall is not doing justice to the usability of these lists : suggestions that are different from

7. Dutch news magazines, science news, documentaries, educational programs and archival video, see `http://www.academia.nl`

the manually assigned keywords are sometimes relevant. This can be compared with the typical values of inter-cataloguer consistency, which range from 13% to 77% (with an average of 44%) when a controlled vocabulary is used (Leininger, 2000) : two cataloguers would, to a large extend, not agree with each other's annotations. The topology of disagreement shows that a portion of these differences are small semantic differences. To reduce the shortcomings of an evaluation based on a strict string-based comparison, we also performed a second type of evaluation according to the research of Medelyan and Witten (Medelyan & Witten, 2005). Their adapted measure of precision and recall takes a semantic distance into account : suggested keywords (extracted automatically) and manually assigned ones that have one thesaurus relationship between them are also counted as correct. Hence, if the keyword selected by the cataloguer is *Salsa* and we propose *Latin American Music*, we consider the suggestion as *semantically correct*.

Hence we have used the following evaluation scheme : (1) classic and semantic precision and recall of the ASR-based keywords suggestion against manually assigned ones (for 110 items), (2) comparison of these precision and recall measures with the ones of the keywords suggestions extracted from online websites related to the same TV programs (for 13 items only).

4.3 Results

Strict Precision and Recall of the ASR and Context Documents-based Keywords Suggestions Table 1 shows the "strict" precision, recall and F-score of the keywords suggestions, both the keywords generated from ASR transcripts and from context documents. Examples of such keywords (translated from Dutch) are :

– Keywords from cataloguers : Plane Industry, Planes, Aviation ;
– Keywords from ASR : Men, Passports, Cities, Aiports, Lakes, Airplanes, Young men, Color [8] ;
– Keywords from Context documents : Factories, Interview, Fighters (Planes), Hunters, Armed forces, Models, Airplanes, Seas, People,, Documentaries, Bees, Bombardments, Cameras [9].

The measures are evaluated against the cataloguer's manually assigned keywords. The table displays the general trends of the Figure 1 [10] : the values for the ranks 1, 3, 5 and 10. The scores are quite low, but the results from the ASR files are comparable to the ones from the context documents. These two sources for generating keywords have opposite shortcomings : the ASR files are very long, hence jeopardizing the precision of the results, whereas the context documents are very short, generating only few keywords [11] and leading to a low recall.

Besides these three measures, we also evaluated the *proportion* of retrieved keywords : the number of keywords that should be proposed (*i.e.* the ones selected by the professional cataloguers) and that indeed figure in the list of propositions. This time, it is much higher for the ASR set than those based on the context documents. This proportion ranges, when not null, from 25% to 80% and is higher than the Apolda-based set for all but two cases. The Apolda-based set ranges from 12.5 to 33%, except for one case where the annotation consists of only one keyword, which happens to be part of the context document, and is null in 5 of the 13 cases. These numbers seem to hint that the cataloguers take more inspiration in what is said in the TV program to annotate than in the texts provided by broadcasters for describing them.

8. Non exhaustive list.
9. Non exhaustive list.
10. The Figure represents the values of the f-measure in percentages, for ranks one to ten.
11. A text of a few lines only has few chances to contain strings that refer to GTAA terms.

ASR	@1	@3	@5	@10
precision	0.22	0.17	0.14	0.09
recall	0.07	0.13	0.18	0.24
F-score	0.11	0.15	0.16	0.14
Context documents	@1	@3	@5	@10
precision	0.23	0.18	0.14	0.13
recall	0.02	0.06	0.08	0.22
F-score	0.04	0.09	0.10	0.16

TABLE 1 – Classical evaluation of ASR and Context Documents based keywords, over 110 documents

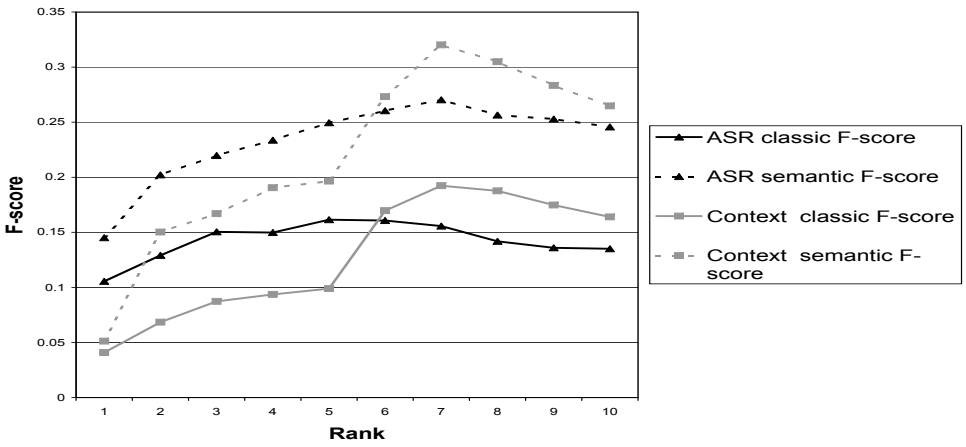

FIGURE 1 – The f-measure for the ASR-based keywords and context document-based keywords, over 110 documents

Precision and Recall for the Two Sets of Extracted Keywords with Semantic Distance Table 2 shows that we improve the results when we allow a distance of one thesaurus relationship between the extracted keyword and the manual annotation of reference. If we look for example at the precision and recall @5 of the ASR-based keywords, we see that on average 1 in 5 suggestions is semantically correct and that we retrieve one third of the keywords in the catalogue description. We get similar results for the set generated from the context documents, despite the fact that the context documents on average are less than a tenth of the ASR text in length.

These results are worse than the ones of previous experiments on context documents (Malaisé *et al.*, 2007) ; the reason for this low performance is the fact that the context documents taken into account here are short, hence giving few hits and few context to each other to generate a good ranking. The ASR have the inverse shortcoming : the files are very long, hence, the ranked precision and recall achieve about the same score as the context documents-based keywords suggestion. In the latter case, a more fine grained ranking algorithm would possibly give better results. But in either case, it is quite interesting to see that in this realistic setting (for most archived programs the available context documents are short), the ASR gives results of equivalent quality as the context documents. ASR is also an interesting alternative for the generation of keywords, even though it is error prone, because it enables to generate annotations from the tv programs' content itself, not on context information which might not be archived or accessible anymore.

ASR	@1	@3	@5	@10
semantic precision	0.25	0.23	0.21	0.17
semantic recall	0.10	0.21	0.31	0.44
semantic F-score	0.15	0.22	0.25	0.25
Context documents	@1	@3	@5	@10
semantic precision	0.23	0.28	0.25	0.20
semantic recall	0.03	0.12	0.16	0.39
semantic F-score	0.05	0.17	0.20	0.26

TABLE 2 – Semantic evaluation of ASR and Context Documents based keywords, over 13 documents

Discussion The fact that the precision and recall figures are equivalent for the two sets is due to the fact that the ASR is producing long lists of possible keywords that contain a small number of correct ones, whereas this proportion is higher for the context documents, although they generate only small lists. In the case of ASR, the challenge lies in finding the most relevant (and correct) suggestions at the top of the list, as the ASR-based keyword suggestions contains a higher number of keywords from the reference set (made by the cataloguers). This observation stresses the fact that the ranking algorithm is the crucial part of our keyword suggestion pipeline, especially in the case of using ASR as a basis for keyword suggestions.

5 Conclusion

In this paper we evaluated the suitability of automatic speech recognition transcripts as a textual source for automatic keyword suggestion. This approach has the benefit that it can also be used for keyword suggestion when audiovisual documents cannot easily be connected to collateral textual context, e.g., on the web. A possible caveat when using speech recognition technology could be that transcription accuracy varies a lot across collections, ranging form $10-20\%$ Word Error Rate in the broadcast news domain to above 50% in more difficult domains (heterogeneous spontaneous speech, historical recordings). Especially for the latter category, transcript error rates might simply be too high to be useful for keyword suggestion.

To get an idea of the value of ASR transcripts for our purpose we compared its use with that of available text documents associated with the broadcasts. The analysis of context documents showed good results in a previous study on documentary programs. For these documentaries, the context documents were broadly available, each containing also a lot of relevant information for our purpose. However, we have seen that this might not always be the case : for the Academia collection we were able to find only 13 context documents for our corpus of 110 broadcasts. Furthermore, the average length of these context documents was much shorter than for the study on documentary programs. It shows that in practice it is hard to find textual data usable for the suggestion of keywords for audiovisual documents.

The present experiment showed that the actual value of the context documents for keyword suggestion was not better than ASR output. Although the context documents attain similar performance with much less information, the total amount of good suggestions contained by the ASR lists is larger. The problem is that the number of wrong suggestions is also larger, so the ratio's in terms of precision, recall and F-score are the same for the context documents and the

ASR output. This means that there is ground for improvement in the form of better ranking algorithms. This is an interesting option for further research. Another way of improvement could be optimization of the ASR engine itself. The system setup used for these experiments was not optimized towards recognition of the GTAA thesaurus terms.

There is also ground for improvement for the context documents, but the possible gain is smaller as the length of their suggestion lists is smaller : the recall of their total suggestion list is lower than for the ASR. In general we can conclude that ASR transcripts seem to be a useful alternative for keyword suggestion. Even though the precision and recall numbers are generally not very high, they still seem valuable, especially when no other text sources are available, which actually may be quite common. For situations where long context documents can be found, keyword extraction and ranking performed on these context documents should still be preferred.

Acknowledgements

This paper is based on research funded by the NWO program CATCH (`http://www.nwo.nl/catch`) and the EU FP7 project Living Web Archives (`http://liwa-project.eu`).

Références

BAAYEN R. H., PIEPENBROCK R. & GULIKERS L. (1995). *The CELEX Lexical Database*. Philadelphia, PA : Linguistic Data Consortium, University of Pennsylvania, (release 2) [cd-rom] edition.

BRUGMAN H., MALAISÉ V., GAZENDAM L. & SCHREIBER G. (2008). The documentalist support system : a web-services based tool for semantic annotation and browsing. Semantic Web Challenge track of the International Semantic Web Conference 2008 (ISWC 2008).

BYRNE W., D.DOERMANN, FRANZ M., GUSTMAN S., HAJIC J., OARD D., PICHENY M., PSUTKA J., RAMABHADRAN B., SOERGEL D., WARD T. & ZHU W.-J. (2004). Automatic recognition of spontaneous speech for access to multilingual oral history archives. *IEEE Transactions on Speech and Audio Processing*, **12**(4), 420 – 435.

CIRAVEGNA F. & WILKS Y. (2003). *Annotations for the Semantic Web*, volume 1, chapter Designing Adaptive Information Extraction for the Semantic Web in Amilcare. IOS press.

GAZENDAM L., MALAISÉ V., DE JONG A., WARTENA C., BRUGMAN H. & SCHREIBER G. (2009). Automatic annotation suggestions for audiovisual archives : Evaluation aspects. *Interdisciplinary Science Reviews*, p. In press.

GAZENDAM L., MALAISÉ V., SCHREIBER G. & BRUGMAN H. (2006). Deriving semantic annotations of an audiovisual program from contextual texts. In *Proceedings of First International workshop on Semantic Web Annotations for Multimedia (SWAMM 2006)*.

HANSEN J., HUANG R., ZHOU B., DEADLE M., DELLER J., GURIJALA A. R., KURIMO M. & ANGKITITRAKUL P. (2005). Speechfind : Advances in spoken document retrieval for a national gallery of the spoken word. *IEEE Transactions on Speech and Audio Processing*, **13**(5), 712–730.

HUIJBREGTS M. (2008). *Segmentation, Diarization and Speech Transcription : surprise data unraveled*. PhD thesis, University of Twente.

HUIJBREGTS M., ORDELMAN R. & DE JONG F. (2007). Annotation of heterogeneous multimedia content using automatic speech recognition. In *Proceedings of SAMT 2007*, Genova, Italy.

KAHAN J. & KOIVUNEN M.-R. (2001). Annotea : an open rdf infrastructure for shared web annotations. In *World Wide Web*, p. 623–632.

KIRYAKOV A., POPOV B., TERZIEV I., MANOV D. & OGNYANOFF D. (2005). Semantic annotation, indexing, and retrieval. *Journal of Web Semantics*, **2**(1), 49–79.

KOHLER J., LARSON M., DE JONG F., KRAAIJ W. & ORDELMAN R. (2008). Spoken content retrieval : Searching spontaneous conversational speech. *ACM SIGIR Forum*, **42**(2), 67–76.

LEININGER K. (2000). Inter-indexer consistency in psycinfo. *Journal of Librarianship and Information Science*, **32**(1), 4–8.

LESPINASSE K. & BACHIMONT B. (2001). Is peritext a key for audiovisual documents ? the use of texts describing television programs to assist indexing. In *CICLing '01 : Proceedings of the Second International Conference on Computational Linguistics and Intelligent Text Processing*, p. 505–506, London, UK : Springer-Verlag.

MALAISÉ V., GAZENDAM L. & BRUGMAN H. (2007). Disambiguating automatic semantic annotation based on a thesaurus structure. In *14e conference sur le Traitement Automatique des Langues Naturelles (TALN)*.

MEDELYAN O. & WITTEN I. H. (2005). Thesaurus-based index term extraction for agricultural documents. In *Proceedings of the 6th Agricultural Ontology Service (AOS) workshop at EFITA/WCCA*.

SMEULDERS A. W. M., WORRING M., SANTINI S., GUPTA A. & JAIN R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(12), 1349–1380.