2014-06-15

Semantic Ontologies for Multimedia Indexing (SOMI): Application in the e-Library Domain

Issam Bendib
University of Tébessa, Algeria, bendibissam@yahoo.fr

Mohamed Ridda Laouar
University of Tébessa, Algeria, ridda_laouar@yahoo.fr

See next page for additional authors

Original Publication Citation
Library Hi Tech 32 (2): 206-218

Authors
Issam Bendib, Mohamed Ridda Laouar, Richard Hacken, and Mathew Miles

Semantic Ontologies for Multimedia Indexing (SOMI): Their application in the e-Library Domain

*BENDIB Issam, *LAOUAR M. Ridda, **Richard Hacken and ***Mathew Miles

*LAMIS Laboratory, University of Tebessa, Algeria
**Brigham Young University, Provo, Utah, USA
***Brigham Young University-Idaho, Rexburg, Idaho, USA.

1 Introduction

The development of techniques for intelligent and reliable retrieval of all kinds of data formats has become increasingly important as multimedia and print resources proliferate at ever increasing rates. Keyword search techniques used by current library systems are insufficient for retrieving multimedia documents and are seen as inferior to the more sophisticated textual searching offered by Google.

One way to improve searches of multimedia documents is to integrate semantic indexing within domain ontologies. To understand this concept we must define its component parts.

Indexing

Librarians are familiar with the concept of indexing, i.e., the creation and organization of access points within metadata in a way designed to facilitate fast and accurate information retrieval.

Semantics and the Semantic Web

Semantics is the study of meaning, investigating the relationship between *signifiers* like words and symbols and what they stand for, their *denotation*. The Web is evolving over time into a Semantic Web whose aim is to automate the process of discovering and collocating units of common meaning previously scattered across the Web. To accomplish this goal, the Semantic Web has been steadily adopting common standards that allow for the sharing and integrating of data. Though much progress has been made, it is clear that conventional tools fall far short in accomplishing this goal due to the sheer volume of Web content. The crux of the problem is that traditional indexing requires too much human interaction for metadata creation and integration. New data is being created at such a pace that there simply aren't enough humans to keep up. We see this problem daily in our libraries. Though discovery tools like Ex Libris Primo, Serials Solutions Summon, and EBSCO Discovery Service attempt to aggregate content from a variety of sources, they fail in terms of pulling data from all the sources that exist – including the free Web. The major interoperability standards for the Semantic Web have been created, *e.g.,* XML, Resource Description Framework (RDF), and Web Ontology Language (OWL). Creating standards is one thing but building new systems and converting legacy systems and data to implement those standards is quite another.

Domain Ontologies

In the computer realm, domain ontology is a conceptual model of the elements of a system, i.e., a model relating to the kinds of objects, properties, etc. throughout the system rather than to the objects themselves. Domain ontologies encompass a scope of related models and their associations with each other. For example, a library system (domain) is composed conceptually of a variety of objects and their relationships such as book, item, and patron. The idea of a book in a library system along with its attributes (title, subject, author etc.) is part of the conceptual model. Further modeling of the book would include the idea that a book title could have multiple copies, which could be checked out by different people.

The term Semantic Ontologies for Multimedia Indexing (SOMI) describes the conceptual models and associations between models that facilitate the automation of the indexing process for multimedia. Current indexing models are not capable, in practice, of automating the index process to this extent.

The elements of the SOMI approach combine three intrinsic multimedia properties:

- The media property (ontological element) allows SOMI to assemble the constituent media objects,
- The semantic property allows SOMI to further encapsulate associations between the media objects,
- The functional property (ontological element) allows SOMI to define phonic indexing techniques and to invoke semantic descriptions within applications.

Based on these properties, we show in sections 3.1 and 3.2 how SOMI's automated index searching overcomes the problem of manual annotation. In section 3.3, we represent the indexes with dense units in a model (opting to use sub-word-based probability modeling). The last section of the architectural overview, section 3.4, illustrates how SOMI's ontological insights can be used to describe semantic concepts for query refinement within the advantages of word confusion networks (WCNs), which use algorithms to process phonic-to-text packets of words or sub-words for prioritization via probability processing, thus rapidly improving the quality of the indexing process.

SOMI ontologies constitute flexible, versionable units that can be concurrently modified and authored within a distributed collaborative environment.

2 An Overview of Previous Works

There are two basic strategies to integrate automated speech recognition for indexing. The first strategy is to use the speech recognizer to produce a transcript as close as possible to what was said in the speech document [7]. The second strategy is to have the recognizer output information useful for the retrieval process, but not necessarily meaningful to human users. Under this second strategy we distinguish two categories. The first category employs the use of keyword spotting, a recognition mode in which no transcripts are generated and the task of the recognizer is to extract key elements only. The second category attempts to leverage the N-best list or other representations of the N-most-likely hypotheses generated by the recognizer [5].

To explain the N-best list, there are many scenarios that arise where it may be difficult to disambiguate a string of spoken text. For example, assume the speaker uses in discourse a word that phonetically resembles "Cosco" or "Costco." Is the word referring to the former, a shipping company, or the latter, a large retailer? Instead of returning a single result based on anterior and posterior words (context), the N-best list is returned which is a list of multiple likely possibilities. Some processing is then applied to this list and it is re-ordered based on *probability* of correctness, allowing the recognizer to more accurately interpret the meaning of spoken text. The problem with N-best lists is that they require significantly more processing power, which degrades the speed of the recognizer. Therefore, N-best lists are not applied to all speech by the recognizer. There is a balance of processing speed to speech recognition accuracy that must be determined by the programmer in order to find the optimal threshold. If the N-best list threshold is set too high, the recognition software will be too slow. If it is set too low, the accuracy will degrade significantly.

In this paper we are interested only in the second strategy, since the first requires massive computations, proving impossible when working with minute phonic variations within huge volumes of resources (e.g., large digital libraries). Navigation technologies and information retrieval are rapidly evolving to meet the diverse needs of multimedia applications. However, the enormous masses of data make it difficult for experts to automate their indexing. New challenges call for appropriate modeling toward the development of applications.

A number of different methods for spoken document retrieval (SDR) using lattices have been proposed. For instance, Siegler used word lattices instead of phone lattices (phonic graphs) as the basis of retrieval, and generalized the *tf-idf* (term frequency to inverse document frequency) ratio to allow for uncertainty in word counts. Chelba and Acero preprocessed lattices into more compact position-specific posterior lattices (PSPL), and computed an aggregate score for each document based on the posterior probability of edges and the proximity of search terms in the document. Mamou et al. converted each lattice into a word confusion network and estimated the inverse document frequency (*idf*) of each word *t* as the ratio of the total number of words in the document collection to the total number of occurrences of *t*. [3].

Jones et al. describe a system that combines a large-vocabulary continuous speech recognition (LVCSR) system with a phone-lattice word spotter (WS) for retrieval of voice and video mail messages [4]. Srinivasan and Petkovic introduce a method for phonetic retrieval based on the probabilistic formulation of term weighting using phone confusion data [11]. Logan et al. compare three indexing methods based on words, syllable-like particles and phonic units to study the problem of out of vocabulary (OOV) queries in audio indexing systems. They have given an alternate approach to the OOV query problem by expanding query words into in-vocabulary phrases while taking acoustic confusability and language model scores into account [6].

Retrieval performance increases by extracting alternative hypotheses from the speech recognizer in addition to the most probable (one-best) candidate. A lattice is a graph containing a number of most probable hypotheses considered by the recognizer and can be used as a source for extracting additional terms. A more compact representation for the hypotheses is the word confusion network (WCN), which offers a convenient representation of competing terms along with the posterior probability for each term. Mamou et al. have shown improvements of SDR performance in low accuracy conditions by indexing and weighting terms in confusion networks based on their probability and rank among competitors [8].

3 Architectural Overview

At present digital libraries are growing at unprecedented rates, especially in terms of unstructured documents or multimedia. Harvesting from the audio component of these unstructured resources requires great efforts to facilitate indexing and access, so the search for a multidisciplinary approach to the problem is a major opportunity. We find that multimedia indexing is more a conceptual problem than it is a technical one. Indexing is traditionally based on two clearly distinguishable steps:

- A step of conceptual analysis: the content is analyzed and interpreted by a librarian to identify key concepts that characterize it;

- A documentary reformulation step: the conceptual analysis allows the librarian to reformulate the content in a form suitable for manipulation.

Our contribution to this discussion is to propose an approach that combines semantic ontological indexing techniques for multimedia documents (SOMI) in the preliminary, conceptual step in order to exploit new automated technologies of information and knowledge. The integration of semantics into document analysis allows us to have solid and reliable automated indexing. The semantic web and domain ontologies encourage us to work with multimedia resource content in this manner. Given the barriers to large-scale manual annotation, our contribution to the automation of metadata retrieval is presented in Fig. 1.

3.1 Indexing with Large Vocabulary Recognition and Keyword Spotting

To search indexes, we execute a large vocabulary speech recognition system on segments extracted from the target multimedia resource. Using the language model, we then retain the frequent terms (*tf*). In order to overcome the problem of locality (common term for one segment only but not for the entire document), we repeat this task using multiple segments for the same resources.

In speech recognition, given an observed acoustic signal O, the goal is to find the sentence $\hat{S}$ that most likely gave rise to O. Using the noisy channel metaphor,

$$\hat{S} \doteq \arg max_s P(S/O) = \arg max_s P(S)P(O/S)$$

Here the language model is the distribution over sentences P(S*)*.

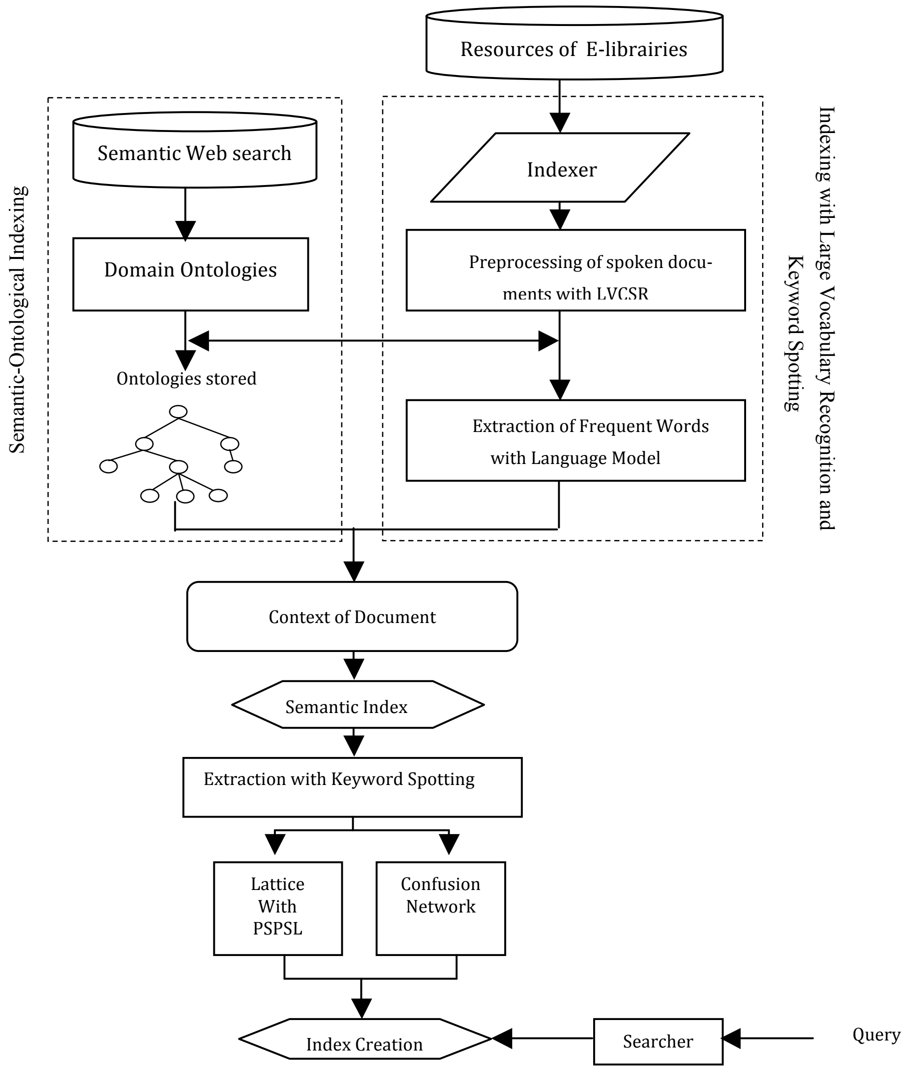

Figure 1: Architecture of SOMI

The acoustic models are based on the state-of-the-art hidden Markov models (HMM) architecture using standard three-state left-to-right models with a mixture of multiple Gaussians in each state. The language models (query likelihood models) are trigram models. The pronunciation dictionaries contain few alternative pronunciations. The automated speech recognition (ASR) system is a single pass system. The recognition networks are represented as weighted finite state machines (FSMs). The output of the ASR system is also represented as an FSM and may be in the form of a best-hypothesis string or a lattice of alternate hypotheses. The labels on the arcs of the FSM may be words or phonic units, and the conversion between the two can easily be done using FSM composition. Timing information can also be present in the output. We recap; the following algorithm models this step:

Algorithm 1
$U_0 = \phi$
While each segment k
    $U_k = \phi$
    Run LVCSR
    If term frequency>=Threshold then Insert term to U_k
    $U = U_k \cap U_{k-1}$
End

3.2 Semantic-Ontological Indexing

The semantic resources (thesauri, ontologies, etc.) offer a significant contribution to the treatment of multimedia documents. Their use in information retrieval (IR) can intervene during the research phase or during the indexing phase [9][10]. The indexing phase forms an initial access to document structure that will facilitate the retrieval phase. The more the indexing phase is involved, the easier the retrieval phase will be.

Documents can be indexed by a group of concepts, such as when a document discusses the concepts A and B, but we do not know of any relationship between them. Therefore, we attribute a semantic description to each document where concepts are likewise represented with their semantic relationships. This representation gives a great power of expression but can decrease the quality of processing. The careful construction of semantic descriptions associated with each document is not an easy task.

The ambiguity of words (e.g., of homonyms, homophones and polysemes) poses particular problems to automatic indexing, where technically literal and general language processing is used. This technique does not completely solve the problem, and the result must always be a compromise between the finesse and complexity of the treatment systems. In this context, research has shown the need to index both by concepts (*i.e.*, word meanings or semantics) and by words recognized in the document. Indexing documents by concepts alone may be misleading because the disambiguation techniques are not completely reliable, while relying solely on words may cause a loss of information.

In our approach, we index terms irrespective of whether or not they are connected to a domain-specific ontology. The semantic links are a plus, but they can also be found as terms unconnected to the ontology.

Concept indexing is defined as the process of identifying instances (entities) and abstract ideas (concepts) within the audio output of a multimedia document and linking the words and phrases in that document to ontological concepts. A concept index is a machine-understandable index of entities and concepts contained in document collections. An entity is an identifiable and discrete instance existing in a multimedia document. A concept is an abstract or general idea inferred or derived from specific instances.

The main assumption underlying concept indexing is that the information conveyed in a document can be analyzed in terms of the entities and concepts that the text contains. This approach involves three steps:

- extracting entities from unstructured document-based content using lexical models and rules,
- identifying concepts and adding ontological tags to them using semantic rules, and
- merging entity and concept information into a concept index.

To this end, we propose a phase that exploits the semantic information provided by a Semantic Web search. The semantic indexing process is responsible for the import of specific domain ontologies (if needed), the definition of metadata describing the proceedings and their storage in the system. Thus, we manage the content of multimedia documents by specific domain ontologies as assisted by semantics.

3.3 Index Modeling

After the detection and selection of frequent terms, these will then be used as index terms. We present these data whenever possible by using multi-level (word and subword) information to improve the performance for both out-of-vocabulary (OOV) and in-vocabulary (IV) words. It has been verified that the phonic-based indexing method is especially effective for OOV keywords. Generally, though, this method yields lower precision for IV queries than does word-based indexing. The benefit of combining word and phonic hypotheses has been shown in recent works [13]. The combination of phonics with word confusion networks (WCNs) has been reported as effective in achieving high performance retrieval for both IV and OOV queries.

Phone lattice scanning (PLS) is a method of indexing audio files from phonetic transcriptions using acoustic-phonetic decoding [1]. There are studies using this method for speech-file indexing of unlimited vocabulary size. This approach consists of generating a phonic lattice in differentiated time for each audio file using a modified version of the Viterbi algorithm [2]. The detection of keywords is performed by dynamic comparison between the search word and phonic sequences in the lattice. The goal of Position Specific Posterior Lattices (PSPL) is to calculate the posterior probability *prob* of a word *W* at a specific position *pos* in a lattice for a spoken segment *d* as a tuple *(W, d, pos, prob)*. Such information is actually hidden in the lattice L of *d* since in each path of L we clearly know each word's position. Since it is very likely that more than one path includes the same word in the same position, we need to aggregate over all possible paths in a lattice that include a given word at a given position.

A variation of the standard forward-backward algorithm can be employed for this computation. The forward probability mass α(*W, t*) accumulated up to a given time *t* at the last word *W* needs to be split according to the length *l* as measured by the number of words:

$$\alpha(w,t,l) \doteq \sum_{\substack{\pi\text{:a partial path ends at time } t,\\ \text{has last word } W\text{ ,and inludes}\\ l \text{ ssubword units}}} P(\pi)$$

where π is a partial path in the lattice. The backward probability *β(W, t)* retains the original definition [13].

The elementary forward step in the forward pass can now be carried out as follows:

$$\alpha(W,t,l) = \sum_{w'} \sum_{\substack{t'\text{:}\exists\text{ edge } e\\ \text{starting at}\\ \text{time } t',\text{ end-}\\ \text{ing at time } t,\\ \text{and with word}(e)=W}} \left[\alpha(W',t',l') \cdot P_{AM}(w) \cdot P_{LM}(w)\right]$$

where $P_{AM}(W)$ and $P_{LM}(W)$ denote the acoustic and language model scores of *W* respectively; *e* is a word arc in the lattice and *word(e)* is the word entity of arc *e*.

The position specific posterior probability for the word W being the *l*th word in the lattice is then:

$$P(W,b,b+Sub(W)-1|L) = \sum_{t} \frac{\alpha(W,t,b+Sub(W)-1) \cdot \beta(W,t)}{\beta_{start}} \cdot Adj(W,t)$$

where *βstart* is the sum of all path scores in the lattice, and *Adj(W, t)* consists of some necessary terms for probability adjustment, such as the removal of the duplicated acoustic model scores on *W* and the addition of missing language model scores around *W* [13]. In this context, we regard the tuples (*W, d, pos, prob*) for a specific spoken segment *d* and position *pos* as a cluster, which in turn includes several words along with their posterior probabilities**.**

3.4 Improving the Quality of Indexing via Word Confusion Networks (CSNs)

The confusion network is the most compact structure representing multiple hypotheses while maintaining the order of symbols (phones/words) along the time axis, meaning that the space for the index table can be reduced. In addition, the confusion network essentially has more paths than the original lattice (which has only those paths allowed by the speech recognizer. For indexing, confusion networks offer a convenient source for expanding the transcript with alternative recognition candidates. Confusion networks are more compact than lattices, and they also provide alignment for all the terms in the lattice. With confusion networks, it is easy to rank locally competing terms by their posterior probability and use the information for indexing [12]. Generally, the following steps compose the algorithm for transforming a lattice to a confusion network:

- Compute the posterior probability for all edges in the lattice
- Pruning: remove all edges with posterior probability below some threshold
- Intra-word clustering: merge together edges corresponding to the same word instance and sum their posterior probabilities
- Inter-word clustering: group different words which compete around the same time interval and have similar phonetic properties to form confusion sets

In this usage, let *D* be a document modeled by a confusion network. We use two pieces of information in the confusion network for each occurrence of a term *t* at position *o*: its posterior probability *Pr(t|o,D)* and its rank among competitors *rank(t|o,D)*. Posterior probability tells how confident the recognizer is that the term occurs in the signal at that position. Rank of the term reflects the importance of the term relative to the other alternatives. In retrieval, a term with a higher probability and/or higher rank should be preferred to one with lower values.

The classical vector space model with *tf-idf* weights and cosine distance relevance measure is used for ranking the search results [12]. Normally, term frequency *tf* is the number of times a term occurs in a document. In our case, we need to estimate a value for term frequency based on the posterior probabilities and ranks of each occurrence of the term in the confusion network of a document.

The term frequency is evaluated by summing the posterior probabilities of all of its occurrences in the confusion network. This means that if the recognizer is confident that it has correctly recognized the term at a given location (with a posterior probability close to one), then the term frequency increases by (close to) one as in the case of indexing error-free text documents. Less weight is given to terms with less confidence. Thus, the term frequency of a term *t* in a document *D*, *tf(t,D)* is defined:

$$tf(t,D) = \sum_{i=1}^{|occ(t,D)|} \Pr(t|o_i,D)$$

The inverse document frequency *idf* indicates the relative importance of a term in the corpus. Traditionally, *idf* is a function of the number of documents in the collection in which the term occurs. In our case, we count the number of confusion networks where the term occurs at any position. In other words, term occurrence *o(t,D)* is estimated by:

$$o(t,D) = \begin{Bmatrix} 1, & if\ tf(t,D) > 0 \\ 0, & otherwise \end{Bmatrix}$$

Now, the inverse document frequency for a term *t* is

$$idf(t) = \log(N / \sum_{D} o(t, D))$$

where *N* is the number of documents in the collection.

In the equation for *o(t,D)*, the value of *tf(t,D)* could also be applied as a threshold by using a value greater than zero to eliminate the effect of terms with low estimated frequency.

4 EXPERIMENTS

4.1 Evaluation Techniques

We use the standard word error rate (WER) as our metric to evaluate system performance. For evaluating indexing system performance we use precision and recall with respect to manual transcriptions. Let *Correct(q)* be the number of times the query *q* is found correctly, *Answer(q)* the number of answers to the query *q*, and *Reference(q)* the number of times *q* is found in the reference.

$$Precision(q) = \frac{Correct(q)}{Answer(q)}$$

$$Recall(q) = \frac{Correct(q)}{Reference(q)}$$

In addition to individual precision-recall values we also compute the *F*-measure defined as

$$F = \frac{2 * Precision * Recall}{Precision + Recall}$$

And report the maximum *F*-measure (*maxF*) to summarize the information in a precision-recall curve.

4.2 Experimental Steps

We will perform several steps in our approach. In the first we execute a speech recognition system and compute the *N*-best result. A posterior probability marks the maximum for the top *N* results. Generally, the top one is the result of the one-best match. Here, letting W_1 and W_2 range over the *N*-best hypothesized output by a speech recognizer, the real result *Wc* (center hypothesis) is found as follows:

$$W_c = arg \min_{i=1,N} \sum_{k=1}^{N} P(W_1^{(k)} | A) WE(W_2^{(i)}, W_1^{(k)})$$

where WE(…) signifies the edit distance of two different strings.

Then, for the semantic treatment of the ontology, a similarity measure based on the informational content of a concept is used. This measure reflects the relevance of a concept in the corpus, taking into account the frequency of its appearance in the corpus as well as the frequency of appearance of concepts that subsume it. We say that a concept *C1* subsumes concept *C2* if *C2* is more specific than *C1*. More precisely, the information content is calculated as per the following formula:

$$CI(c) = -\log\ \left(P(c)\right)$$

where *P(c)* is the probability of finding an instance of concept c. These probabilities are calculated by frequency *(c) / N* where *N* is the total number of concepts [14].

We subsequently define the semantic similarity between two concepts by the amount of information they share. This shared information is equal to the information content of the smallest generalization (*SG*), the most specific concept that subsumes the two concepts in the ontology [14].

$$Sim(c1, c2) = CI(SG(c1, c2))$$

During the next phase we generate the word confusion network using the output of the previous step – which we will prune at various percentiles – and we apply a minimum error training procedure with the goal of weight optimization of the log-linear model. Note that a separate optimization is performed for each *N*-best and *CN* condition.

Finally, we use a standard large vocabulary continuous speech recognition system for generating ASR 3-gram lattices and PSPL lattices. The 3-gram language model used for decoding is trained on a large amount of data, which was selected based on frequency of occurrence found in the training data. The acoustic model is trained on a variety of wide-band speech, a standard clustered tri-phone, three-states-per-phone model.

5 CONCLUSION

The growth and diversity of multimedia resources with audio components in digital libraries render the process of indexing extremely difficult. The cascading volume of documents limits conventional methods of indexing based on human skills alone. As a result, the management of Big Data within the realm of multimedia resources requires us to find an automatic indexing solution. In this context, the integration of semantics is essential to reliable indexing. The advent of the Semantic Web and domain ontologies – not to mention recent advances in speech recognition technology – encourages us to seek hybrid solutions for this problem. However, validation and implementation of this approach within the field of digital libraries remain under development.

6 REFERENCES

1 Chelba, C., Silva, J. and Acero, A. (2007), “Soft indexing of speech content for search in spoken documents”, *Computer Speech and Language*, Vol. 21 No. 3, pp. 458–478.

2 El Meliani, R. and O'Shaughnessy, D. (1995), “Lexical fillers for task-independent-training based keyword spotting and detection of new words”, in *EUROSPEECH, Fourth European Conference on Speech Communication and Technology*, Madrid, September 18-21, 1995, Universität Trier, Trier, pp. 2129-2133.

3 Issam, B. and Ridda, L. (2012), “Approaches for the detection of the keywords in spoken documents application for the field of e-libraries”, in *Neural Information Processing: 19th International Conference on Neural Information Processing (ICONIP’12)*, Doha, Qatar, November 12-15, 2012, Springer Berlin/Heidelberg, pp.196-203.

4 Jones, G.J.F. and Foote, J.T. (1996), “Retrieving spoken documents by combining multiple index sources”, in *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zürich, August 18-22, 1996, ACM Press, New York City, pp. 30–38.

5 Larson, M. (2001), “Sub-word-based language models for speech recognition: implications for spoken document retrieval”, in *Workshop Proceedings on Language Modeling and Information Retrieval*, Pittsburgh, May 31 – June 1, 2001, Carnegie Mellon University, Pittsburgh, pp. 78-82.

6 Logan, B., Moreno, P. and Deshmukh, O. (2002), “Word and sub-word indexing approaches for reducing the effects of OOV queries on spoken audio”, in *Proceedings of the Second International Conference on Human Language Technology Research*, San Diego, March 24-27, 2002, Morgan Kaufmann Publishers, San Francisco, pp. 31-35.

7 Makhoul, J., Kubala, F., Leek, R., Lui, D., Nguyen, L., Schwartz, R and Srivastava, A. (2000). “Speech and language technologies for audio indexing and retrieval,” in *Proceedings of the IEEE* Vol. 88 No. 2, pp. 1338-1353.

8 Mamou, J., D. Carmel, D., Hoory, R. (2006), “Spoken document retrieval from call-center conversations”, in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, August 6-10, 2006, ACM Press, New York City, pp. 51–58.

9 Mitschick, A. (2010), ”Ontology-based indexing and contextualization of multimedia documents for personal information management applications”, *International Journal on Advances in Software*, Vol. 3 No. 1, pp. 31-40.

10 Parul, G. and Sharma, A.K. (2010),“Context based indexing in search engines using ontology”, *International Journal of Computer Application*, Vol. 1 No 14, pp. 49-52.

11 Srinivasan, S. and Petkovic, D. (2000), “Phonetic confusion matrix based spoken document retrieval”, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, July 24-28, 2000, Universität Trier, Trier, pp. 81–87.

12 Turunen, V.T. and Kurimo, M. (2007), “Indexing confusion networks for morph-based spoken document retrieval”, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, July 23-27, 2007, ACM Press, New York, City, pp. 631–638.

13 Wessel, F., Schluter, R., Macherey, K. and Ney, H. (2001), “Confidence measures for large vocabulary continuous speech recognition”, *IEEE Transactions on Speech and Audio Processing*, Vol. 9 No. 3, pp. 288–298.

14 Zargayouna, H. and Salotti, S. (2004), “Mesure de similarité dans une ontologie pour l'indexation sémantique de documents XML“, in *Actes de la conference Ingénierie*

des Connaissances, Lyon, May 5-7, 2004, Nada Matta, Lyon, available at: http://liris.cnrs.fr/~ic04/programme/.