Semantic indexing of continuous speech

Mohamed Labidi#1, Naim Terbeh#2 Mohsen Maraoui*3, Mounir Zrigui#4

Research Laboratory of Technologies of Information and Communication & Electrical Engineering LaTICE (Monastir unit)
Faculty of Science of Monastir, computer science department, Monastir 5000, Tunisia
1labidi8mohamed@gmail.com
2terbehnaim1987@gmail.com
4zrigui.mounir@fsm.rnu.tn

*Computational Mathematics Laboratory, University of Monastir,
Faculty of science of Monastir, Mathematic department, Monastir 5000, Tunisia
3maraoui.mohsen@gmail.com

Abstract— **Speech, as the most important means of communication between human beings, underwent several scientific studies, but until now there remains much to do with this phenomenon. After more than 60 years of research, the representation of speech has not exceeded the representation of the content (standard MPEG7). This paper presents a step towards a standard semantic indexing of continuous speech (indexing by the sense). The index is presented based on the meaning of the elements (Verbs, Names) of the sentences. Also other information that we can extract from speech. The index will be simpler and more meaningful. We tried not to use any external source to represent the meaning, which makes it more difficult and need resources that not in mostly exist. Also it focuses only on the speech, making specific and effective index, and because the speech she is rich enough to have its own standard. This work presents a step towards our goal of a standard semantic indexing continuous speech.**

Keywords: **speech, semantic, sense, standard, indexation, TALN, speech recognition, multimedia.**

I. Introduction

Indexing is designed to facilitate research. This research which it differs according to user needs. These needs are related to the existing information in speech.
The search for the voice information can be limited in the following types:
Search by speakers: This research is used for example to look for words according to their number of speakers. Also search by speakers may be more accurate, if for example the user wants to specify the kind of speaker (female, male, child ...).
Search by tone of voice: The user can also specify the tone of voice (yelling, whispering, speedy pronunciation...)
Search by content delivered: This is similar to text search, where the user is seeking one or more terms (words) in speech.
Search by sense: This research allows the user to search not only the terms (words, phrases ...) only syntactically, but also to seek their senses, the sense of the compositions of words, texts.
Mixed Search: this type of research done in this component one or more searches related to those described above. For example, a user want to find a word that has the meaning X, a conversation between at least two (2) speakers and one of this two speakers blaring.

For the speech we have two generations of Indexers (Indexations standards). The first generation of standards has focused on the speech coding include e.g MPEG-1 standard, MPEG-2 and MPEG-4. The second generation standard is the representation of multimedia content, that is to say what is inside. This generation has appeared with the MPEG-7 standard [1].
MPEG-7 is presented as a standard for the description of multimedia content (video and audio). What interests us here is the representation of the audio content. This description was executed in the following four parts: Audio framework.,Spoken Content DS, Timbre Description and Audio Independent Components.
For the semantic indexing of the text, much work has been carried out. One example is the work of [11] which proposed a statistical approach to index multilingual documents. This approach was compared with other approaches based on language resources and has given results similar to those of linguistic approaches. For linguistic approaches to semantic indexing we can cite the work of [12], [13] and [14] who used UMLS (Unified Medical Language System) to index the ImageCLEFmed collection. Similarly the thesaurus MeSH (Medical Subject Heading) is used in [15] to index TREC documents. In [16] the authors recently proposed a data model for indexation based on a representative reference ontology semantic indexing terms. The proposed model aims to enable real-time indexing following the dynamics of the corpus while ensuring the availability of documents and index.

The list of existing standards used to represent multimedia content in its best standard (MPEG7). But for the speech, we need more than that. The information in the speech signal exceeds the content to other things. Specifically talking about the meaning of the message contained in the speech. This paper describes a semantic indexer speech. Speech is rich in information. The information is grouped together to make sense of the message. This paper uses all possible information we can extract from the speech, to create a semantic index. Index with a new vision of the meaning and also a new representation of meaning.

II. INDEXER SEMANTIC

A. Features of an audio file:

An audio file is a set of sentences spoken by one or more speakers. This pronunciation is not only contains all the phonemes compounds together to build words and sentences. But it also includes the way in which these sentences are pronounced by the (s) speaker (s), such as fast pronunciation, slow, loud, exclamation sound, interrogation... Also the spoken sentences are connected in a fixed order. A succession of sentences, which is not arbitrary, but it is well studied by the human cognitive system to export a set of ideas (messages that have meaning) in the form of words, sentences. It should be remembered that the signal also includes information on the speakers themselves (characteristics of the voice identified relative to the other). So according to what is said, a representation of the meaning of the message in an audio file (signal) must take into account all the informations, and the relationships between them.

B. Indexing System:

The figure below shows our proposal for the indexer.

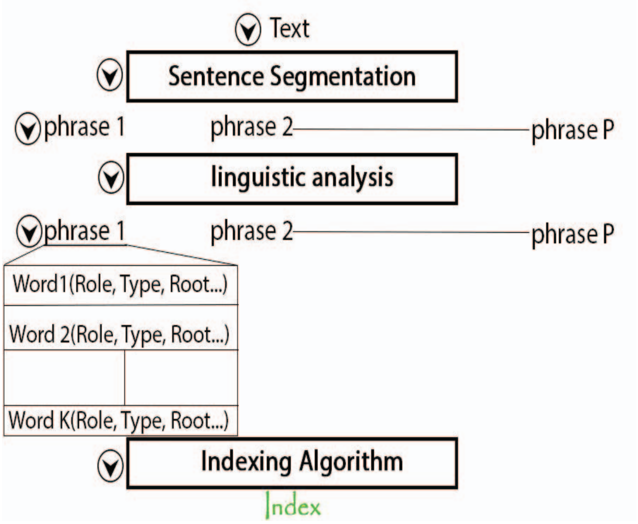

Fig. 1 Steps indexing of text sounded by an automatic speech recognition system.

Each input text will be increased by a set of sentences. Each sentence will undergo a linguistic analysis. This analysis will annotate each word in the sentence with several annotations such as the role in the sentence, the word type, root ... After the annotation, phrases will be given to the indexing algorithm for representing the meaning of the message.

C. Format of the index

This section provides an explanation of each section in our index. The figure below gives more details on the index.

1
Heading
2
Message
&
meaning
3
Summary
of the
Meanings

Fig. 2 Large parts of the index

Each issue has a level in the index. Each level consists of a set of elements that describes the.

1) Heading:

The header has information common between all audio files such as Title, frequency, size, date created, date last modified...

2) Message & Meaning:

This part is to present the message contained in the file and its meaning. This part is extensible it may contain one or more semantic descriptors.
When we say "message", we mean the transcript of what is said. The message text is the result of an automatic speech recognition system.
This part is based on linguistic analysis (morphology, syntax...). This text will be segmented into sentences. Since every complete sentence, presents a correct message (in a sense). Each sentence has an element of our overall message. Then each piece of the puzzle (sentence) will be analyzed separately, then these will be gathered together to form the overall meaning of our message.
Each sentence will have an identifier which is his rank (order number) in the text. We must respect the pronunciation order of sentences to keep the message as it is, consequently keep the meaning as it is. Phrases identifiers beginning of (1) and can't exceed (N), which is the number of sentences in the message. After determining the ID of the sentence, we add the identifier of the speaker of this sentence. The speaker must be among the list of "Speakers". The list of speakers is set aside in a separate file in the index file. It will be presented in a separate paragraph. Each sentence will be analyzed linguistically. After linguistic analysis each word in a sentence will be annotated as is shown in Figure 2.
After this work, it is time to present our method for the representation of meaning.
First we will present the direction and vision for the senses.

<u>The_meaning:</u>
Each phrase pronounced is just a set of objects, statements and actions presented in an order and determine a relationship well to convey a specific meaning. These objects pass from one state to another, to and/or receives the shares.
These three elements will be the basis of our representation of meaning in the speech. In all languages, objects were presented by names, and the statements and actions are presented by verbs. These two families are eternal relationship. This relationship is

direct or indirect by linguistic connectors. These three elements are also in relation with time.

Our vision has led us to create a representation of meaning via the three elements presented (Objects, state, action). A presentation without using external resources. The meaning of a text will be presented by a matrix. This matrix shows the objects in our text, their states in chronological order, and actions also in the same order.

	ID speaker	**ID speaker**	**….**	..	…	**ID speaker**
	State 0	Actions	**State 1**		Actions	**State N**
Name 1		Actor/ Suffering				
Name 2		Actor/ Suffering				
Name L		Actor/ Suffering				

Table 1 : Message meaning

Names (rows) having the set of objects which has in the text (The names). Each object has an initial state (state 0), actually it is the state verbs. After the state we found all the actions (action verbs) who entered directly after or simultaneously with the state above those actions. Each box (Actions / name) will contain either the word "Actor", which means that it's an actor or the word "Suffering" to say that the object acted upon. After we find the actions of the new states. These statements include the statements of objects after the completion of previous actions. This cycle repeats until the end of the action in the text which leads us to a set of final states for objects.

3) Summary of the meanings:

This section summarizes the meaning of the message. This gives each item a probability value has its percentage relative to the other objects. And for each object is given their states with their probabilities of existence.

III. Conclusion

This paper describes a semantic indexer continuous speech. Our goal is to succeed in creating a standard semantic indexing of speech. Only for speech, since it is rich information that come together to form the meaning of the message. It remains a great project that will be covered in future publications, and to start with this paper.

REFERENCES

[1] J.P. Haton, C. Cerisara, D. Fohr, Y. Laprie, K. Smaïli *: Reconnaissance automatique de la parole : Du signal à son interprétation.* 1nd ed., Paris, France : Dunod, 2006.

[2] M.W. Berry and M. Castellanos,. *Survey of Text Mining II.* London : Springer, 2008.

[3] B. Delezoide. '*'Modèles d'indexation multimedia pour la description automatique de films de cinéma*''. PhD thesis, Paris France, university VI Pierre et Marie Curie, 2006.

[4] M .Grgic, K. Delac, and M. Ghanbari,. "Recent Advances in Multimedia Signal Processing and Communications", Studies in Computational Intelligenc, vol. 231 : Springer Berlin – Heidelberg, 2010.

[5] B. Haidar. '*'Services d'Indexation Multimedia Distribues*''. PhD thesis, Paul Sabatier university,Toulouse France, IRIT, 2005.

[6] Y. Hayashi, K. Ohtsuki, K. Bessho, O. Mizuno, Y. Matsuo, S. Matsunaga, M. Hayashi, T. Hasegawa, and N. Ikeda. ''Speech-based and video-supported indexing of multimedia broadcast news". In SIGIR, pp. 441–442. ACM, 2003.

[7] J. Köhler, S. Philippi, M. Specht, and A. Rüegg. ''Ontology based text indexing and querying for the semantic web''. Knowl.-Based Syst., 19(8) :, pp.744–754, 2006.

[8] S. R. Subramanya and A. Youssef. "Wavelet-based indexing of audio data in audio/multimedia databases". Multi-Media Database Management Systems, pp. 46–53, : Dayton, OH 1998.

[9] M. Bendris . '*'Indexation audio visuelle des personnes dans un contexte de television*''. Phd thesis, TELECOM ParisTech, Paris France : 2011.

[10] G. Peeters. '*'Indexation automatique de contenus audio musicaux*''. HDR, University of Pierre et Marie Curie,2013.

[11] F. Harrathi, C. Roussey, L. Maisonasse, S. Calabertto. ''Vers une approche statistique pour l'indexation sémantique des documents multilingues''. In INFORSID, 2010, p. 127 - p. 143.

[12] M. Loïc., G. Eric., J.P. Chevallet.,''Combinaison d'analyses sémantiques pour la recherche d'information médicale '', In INFORSID'2009, Toulouse.

[13] E . Gaussier, L. Maisonnasse, and J.P . Chevallet.,''Multiplying concept sources for graph modeling'', In CLEF 2007, LNCS 5152 proceedings.

[14] C. Lacoste, J.P. Chevallet , J.H. Lim , X. Wei, D.Raccoceanu, D.L.T. Hoang, and F. Uillenemot., ''Ipal knowledge-based medical image retrieval in imageclefmed 2006'', In CLEF 2006 Workshop,p 20-22 .

,

[15] S. Neil, T. Vetle, H. Jie, Y. Clement., ''Knowledgeintensive conceptual retrieval and passage extraction of biomedical literature'', In ACM SIGIR, 2007, p. 655-662.

[16] G. Hubert, J. Mothe, B. Ralalason, B. Ramamonjisoa."Modèle d'indexation dynamique à base d'ontologies''. In CORIA, 2009, p. 169-184.

[17] N. Terbeh, M. Labidi, M. Zrigui. ''Automatic speech correction: A step to speech recognition for people with disabilities'' , In ICTA, , pp.1—6.