

Using Text Segmentation to Enhance the Cluster Hypothesis

Submitted by Emmanuel Lemoine on Mon, 10/06/2014 - 18:25

Titre	Using Text Segmentation to Enhance the Cluster Hypothesis
Type de publication	Communication
Type	Communication avec actes dans un congr�s
Ann�e	2008
Langue	Anglais
Date du colloque	2008
Titre du colloque	13th International Conference, AIMSA 2008
Titre des actes ou de la revue	Artificial Intelligence: Methodology, Systems, and Applications
Volume	5253
Pagination	69 - 82
Auteur	Lamprier, Sylvain [1], Amghar, Tassadit [2], Levrat, Bernard [3], Saubion, Fr�d�ric [4]
Pays	Bulgarie
Editeur	Springer
Ville	Varna
ISBN	978-3-540-85775-4, 978-3-540-85776-1
Mots-cl�s	Artificial Intelligence (incl. Robotics) [5], Cluster Hypothesis [6], Data Mining and Knowledge Discovery [7], Database Management [8], Information Retrieval [9], Information Storage and Retrieval [10], Information Systems Applications (incl. Internet) [11], Text Segmentation [12]
R�sum� en anglais	<p>An alternative way to tackle Information Retrieval, called Passage Retrieval, considers text fragments independently rather than assessing global relevance of documents. In such a context, the fact that relevant information is surrounded by parts of text deviating from the interesting topic does not penalize the document. In this paper, we propose to study the impact of the consideration of these text fragments on a document clustering process. The use of clustering in the field of Information Retrieval is mainly supported by the cluster hypothesis which states that relevant documents tend to be more similar one to each other than to non-relevant documents and hence a clustering process is likely to gather them. Previous experiments have shown that clustering the first retrieved documents as response to a user's query allows the Information Retrieval systems to improve their effectiveness. In the clustering process used in these studies, documents have been considered globally. Nevertheless, the assumption stating that a document can refer to more than one topic/concept may have also impacts on the document clustering process. Considering passages of the retrieved documents separately may allow to create more representative clusters of the addressed topics. Different approaches have been assessed and results show that using text fragments in the clustering process may turn out to be actually relevant.</p>
Notes	Date du colloque : 09/2008

URL de la notice <http://okina.univ-angers.fr/publications/ua4472> [13]
DOI [10.1007/978-3-540-85776-1_7](https://doi.org/10.1007/978-3-540-85776-1_7) [14]
Lien vers le document en ligne http://dx.doi.org/10.1007/978-3-540-85776-1_7 [14]

Liens

- [1] [http://okina.univ-angers.fr/publications?f\[author\]=7522](http://okina.univ-angers.fr/publications?f[author]=7522)
- [2] <http://okina.univ-angers.fr/t.amghar/publications>
- [3] <http://okina.univ-angers.fr/bernard.levrat/publications>
- [4] <http://okina.univ-angers.fr/frederic.saubion/publications>
- [5] [http://okina.univ-angers.fr/publications?f\[keyword\]=6729](http://okina.univ-angers.fr/publications?f[keyword]=6729)
- [6] [http://okina.univ-angers.fr/publications?f\[keyword\]=8979](http://okina.univ-angers.fr/publications?f[keyword]=8979)
- [7] [http://okina.univ-angers.fr/publications?f\[keyword\]=6001](http://okina.univ-angers.fr/publications?f[keyword]=6001)
- [8] [http://okina.univ-angers.fr/publications?f\[keyword\]=8847](http://okina.univ-angers.fr/publications?f[keyword]=8847)
- [9] [http://okina.univ-angers.fr/publications?f\[keyword\]=8980](http://okina.univ-angers.fr/publications?f[keyword]=8980)
- [10] [http://okina.univ-angers.fr/publications?f\[keyword\]=8849](http://okina.univ-angers.fr/publications?f[keyword]=8849)
- [11] [http://okina.univ-angers.fr/publications?f\[keyword\]=6003](http://okina.univ-angers.fr/publications?f[keyword]=6003)
- [12] [http://okina.univ-angers.fr/publications?f\[keyword\]=8981](http://okina.univ-angers.fr/publications?f[keyword]=8981)
- [13] <http://okina.univ-angers.fr/publications/ua4472>
- [14] http://dx.doi.org/10.1007/978-3-540-85776-1_7

Publié sur *Okina* (<http://okina.univ-angers.fr>)