Toward a More Global and Coherent Segmentation of Texts

Submitted by Emmanuel Lemoine on Mon, 10/06/2014 - 17:48

Titre	Toward a More Global and Coherent Segmentation of Texts
Type de publication	Article de revue
Auteur	Lamprier, Sylvain [1], Amghar, Tassadit [2], Levrat, Bernard [3], Saubion, Frédéric [4]
Editeur	Taylor & Francis
Type	Article scientifique dans une revue à comité de lecture
Année	2008
Langue	Anglais
Date	2008
Numéro	3
Pagination	208 - 234
Volume	22
Titre de la revue	Applied Artificial Intelligence
ISSN	0883-9514
Résumé en anglais	The automatic text segmentation task consists of identifying the most important thematic breaks in a document in order to cut it into homogeneous passages. Text segmentation has motivated a large amount of research. We focus here on the statistical approaches that rely on an analysis of the distribution of the words in the text. Usually, the segmentation of texts is realized sequentially on the basis of very local clues. However, such an approach prevents the consideration of the text in a global way, particularly concerning the granularity degree adopted for the expression of the different topics it addresses. We thus propose here two new segmentation algorithms—ClassStruggle and SegGen—which use criteria rendering global views of texts. ClassStruggle is based on an initial clustering of the sentences of the text, thus allowing the consideration of similarities within a group rather than individually. It relies on the distribution of the occurrences of the members of each class 1 to segment the texts. SegGen proposes to evaluate potential segmentations of the whole text thanks to a genetic algorithm. It attempts to find a solution of segmentation optimizing two criteria, the maximization of the internal cohesion of the segments and the minimization of the similarity between adjacent ones. According to experimental results, both approaches appear to be very competitive compared to existing methods.
URL de la notice	http://okina.univ-angers.fr/publications/ua4282 [5]
DOI	10.1080/08839510701881391 [6]
Lien vers le document	http://dx.doi.org/10.1080/08839510701881391 [6]

Liens

[1] 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/ua4282
[6] http://dx.doi.org/10.1080/08839510701881391

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