

Amélioration de la performance de l'Analyse de la Sémantique Latente pour des corpus de petite taille

Fadoua Ataa-Allah*, Abderrahim El Qadi**
Siham Boulaknadel*, Driss Aboutajdine*

*Université Mohamed V Agdal Faculté des Sciences, GSCM, B.P. 1014, Rabat
{fadoua_01, siham_06}@yahoo.fr
aboutaj@fsr.ac.ma
http://www.fsr.ac.ma/GSCM/

**Université Moulay Ismail ESTM, route d'Agouray, km.5, B.P. 3103 Toulal Meknes
elqadi@estm.ac.ma

Problématique. Améliorer la performance du LSA pour des corpus de petite taille où l'unité textuelle est représentée par des paragraphes.

Solution. Appliquer des schémas de pondération et utiliser des listes de mots vides.

Application et Résultats

Dans l'objectif d'améliorer la performance du LSA pour des corpus de petite taille, nous avons appliqué une multitude de schémas de pondération dans deux cas d'études : le premier où le prétraitement des corpus est simple ; le deuxième où une liste de mots vides est utilisée [1].

<i>SLVM</i>	<i>ALVM</i>	<i>Nom du corpus</i>	<i>SLVM</i>	<i>ALVM</i>
$\text{Log}(tf+1) \times \text{Idf}$	$Tf \times \text{Idf}$	<i>Blanche Neige</i>	0.73	0.68
$Tf \times \text{Idf}$	$\text{Log}(tf+1) \times \text{Idf}$	<i>Cendrillon</i>	0.62	0.58
<i>Tfc</i>	<i>Tfc</i>	<i>La Belle au Bois Dormant</i>	0.71	0.64
<i>Ltc</i>	<i>Ltc</i>	<i>Le Petit Chaperon Rouge</i>	0.38	0.34
$\text{Log}(tf+1)/\text{Entropie de Shannon}$	$\text{Log}(tf+1)/\text{Normal}$	<i>Le Petit Poucet</i>	0.47	0.50

(a)

ALVM : Avec l'utilisation de la liste de mots vides.

SLVM : Sans l'utilisation de la liste de mots vides.

(b)

(a) - L'ordre des plus importants schémas de pondération améliorant la performance du LSA

(b) - La précision maximale par LSA pour un seuil de 0.9

TAB 1 – Tableau des résultats

L'étude menée a montré que l'application d'un schéma de pondération peut influencer positivement ou négativement la performance du LSA, tandis que l'utilisation de la liste des mots vides n'est pas performante.

Références

- (Deerwester et al. 1990) Deerwester S, Dumais S.T., Furnas G.W., Landauer T.K., Hrashman R., Indexing by latent semantic analysis, Journal of the American Society for Information Science, 41(6), pages 391-407, 1990.
(Dumais 1992) Susan T. Dumais, Enhancing Performance in Latent Semantic Indexing (LSI) Retrieval, 1992, Technical Memorandum Tm-ARH-017527, Bellcore.
[1] <http://snowball.tartarus.org/french/stop.txt>.