Intégration de la similarité entre phrases comme critère pour le résumé multi-document

Maâli Mnasri[1, 2] Gaël de Chalendar[1] Olivier Ferret[1]
(1) CEA, LIST, Laboratoire Vision et Ingénierie des Contenus, Gif-sur-Yvette, F-91191, France.
(2) Université Paris-Sud, Université Paris-Saclay, F-91405 Orsay, France.
maali.mnasri@cea.fr, gael.de-chalendar@cea.fr, olivier.ferret@cea.fr

RÉSUMÉ

À la suite des travaux de Gillick & Favre (2009), beaucoup de travaux portant sur le résumé par extraction se sont appuyés sur une modélisation de cette tâche sous la forme de deux contraintes antagonistes : l'une vise à maximiser la couverture du résumé produit par rapport au contenu des textes d'origine tandis que l'autre représente la limite du résumé en termes de taille. Dans cette approche, la notion de redondance n'est prise en compte que de façon implicite. Dans cet article, nous reprenons le cadre défini par Gillick & Favre (2009) mais nous examinons comment et dans quelle mesure la prise en compte explicite de la similarité sémantique des phrases peut améliorer les performances d'un système de résumé multi-document. Nous vérifions cet impact par des évaluations menées sur les corpus DUC 2003 et 2004.

ABSTRACT

Integrating sentence similarity as a constraint for multi-document summarization.

Following Gillick & Favre (2009), a lot of work about extractive summarization has modeled this task by associating two contrary constraints : one aims at maximizing the coverage of the summary with respect to its information content while the other represents its limit size. In this context, the notion of redundancy is only implicitly taken into account. In this article, we extend the framework defined by Gillick & Favre (2009) by examining how and to what extent integrating semantic sentence similarity into a multi-document summarization system can improve its results. We show more precisely the impact of this strategy through evaluations performed on the DUC 2003 and 2004 datasets.

MOTS-CLÉS : résumé automatique, ILP, clustering, similarité sémantique.

KEYWORDS: automatic summarization, ILP, clustering, semantic similarity.

1 Introduction

Les travaux menés dans le cadre du RA par extraction ont vu la proposition d'un grand nombre de critères de sélection de phrases et d'intégration des résultats de cette sélection pour former un résumé. Toutes ces propositions ont plus ou moins explicitement pour objectif de faire un compromis entre le respect d'une contrainte de taille maximale du résumé à produire, la maximisation de son contenu informationnel et la non redondance des informations qu'il véhicule.

Depuis quelques années, les approches abordant la problématique du RA comme un problème d'optimisation de contraintes fondée sur la programmation linéaire en nombres entiers (ILP) ont montré des

résultats intéressants. Cette approche présente l'avantage d'optimiser conjointement plusieurs critères exprimés de façon très déclarative, ce qui en fait un modèle assez flexible. Néanmoins, la déclinaison la plus répandue de ce modèle, incarnée par (Gillick & Favre, 2009), ne prend en compte la contrainte de non-redondance des informations que de façon implicite et s'interdit de ce fait de bénéficier des travaux sur la paraphrase et l'implicature textuelle, problématique particulièrement importante dans le cadre du RA multi-document. Dans cet article, nous examinons ainsi comment intégrer de façon plus explicite la contrainte de non-redondance informationnelle dans le cadre proposé par Gillick & Favre (2009) en nous focalisant sur la similarité sémantique des phrases.

2 Travaux précédents

À l'origine des approches que nous considérons ici, McDonald (2007) a proposé d'exprimer le problème du RA sous la forme d'un problème ILP dont la fonction d'objectif cherche à maximiser le poids des phrases sélectionnées. Ce poids est pénalisé par la redondance avec les phrases déjà incluses dans le résumé. Le modèle intègre en outre la contrainte de la taille maximale du résumé. Ce problème a ensuite été reformulé par Gillick & Favre (2009) en définissant une fonction d'objectif se focalisant sur la maximisation du poids des bigrammes de mots sélectionnés, toujours sous la contrainte de la longueur maximale du résumé. La non-redondance est quant à elle favorisée de façon implicite. Le poids de chaque bigramme n'étant comptabilisé qu'une seule fois dans la fonction d'objectif, indépendamment de son nombre d'occurrences dans le résumé final, cette fonction tend à être d'autant plus élevée qu'un nombre plus large de bigrammes est sélectionné, ce qui conduit aussi à limiter le nombre d'occurrences de chaque bigramme et donc, la redondance.

Par la suite, différents travaux ont proposé des modèles ILP plus élaborés tandis que d'autres ont mis en œuvre des traitements plus ciblés en amont de la phase ILP. Li *et al.* (2011) suggèrent ainsi de regrouper les phrases par aspect, en l'occurrence de nature événementielle (qui, où ...), et de garder un représentant par cluster, constitué par la compression de ses phrases. La partie ILP se charge de sélectionner les phrases maximisant l'inclusion des aspects les plus importants dans le résumé. La performance de cette approche dépasse légèrement celle des baselines classiques. L'approche de Woodsend & Lapata (2012) répartit quant à elle la sélection de phrases sur des modules indépendants. Chacun prend en compte un critère différent (e.g. couverture en bigrammes, position et style des phrases, compression des phrases, etc.). La sortie de ces modules est passée au programme ILP dont l'objectif est de maximiser le score des phrases donné par la contribution des différents modules de sélection de phrases. Des méthodes supervisées ont aussi montré leur efficacité, particulièrement l'estimation de la fréquence des bigrammes dans le résumé par un modèle de régression. Les fréquences prédites sont considérées comme les poids des bigrammes dans le modèle ILP (Li *et al.*, 2013). D'autres travaux (Li *et al.*, 2015) se sont intéressés spécifiquement à la pondération des bigrammes en combinant l'utilisation des critères internes, comme les fréquences et les positions des bigrammes dans les documents, et des ressources externes comme WordNet, Wikipédia et DBpedia. Le problème du compromis entre la performance et l'efficacité a aussi été abordé. Il a ainsi été établi que l'élagage des bigrammes peu fréquents améliore la vitesse de l'optimisation mais se fait aux dépens des scores ROUGE. Une approche approximative par agrégation de plusieurs solutions optimales a été proposée comme solution possible à ce problème (Boudin *et al.*, 2015).

3 Méthode proposée

La méthode que nous proposons repose sur deux étapes principales : une phase de clustering sémantique des phrases des textes à résumer et une phase de sélection des phrases du résumé. Cette dernière étape est réalisée grâce à un modèle ILP auquel nous intégrons l'information sémantique déduite du clustering. Nous détaillons dans ce qui suit les différentes composantes de cette méthode.

3.1 Clustering sémantique

L'objectif du clustering sémantique en amont de la phase de sélection de phrases est de regrouper les phrases apportant la même information, éventuellement exprimée de différentes façons. Chaque cluster sémantique ne doit donc avoir au plus qu'un seul représentant, plus précisément une de ses phrases, dans le résumé final. Cette étape a ainsi une double vocation : d'une part, diminuer le coût de la sélection des phrases par ILP ; d'autre part, éviter la redondance informationnelle au sein du résumé. Du point de vue des modèles s'inscrivant dans la lignée de Gillick & Favre (2009), cette stratégie permet en outre de prendre en compte de façon indirecte les similarités sémantiques entre concepts (bigrammes) et donc de renforcer la contrainte de présence d'au plus une seule occurrence de chaque concept dans le résumé. Cette similarité est par ailleurs contextuelle, puisqu'inscrite dans le cadre de la phrase, ce qui limite aussi la combinatoire des recherches d'équivalence entre concepts.

Algorithme de clustering : Pour réaliser notre clustering sémantique, nous avons mis en place un algorithme de clustering incrémental opérant un compromis entre notre volonté de mettre l'accent sur la similarité entre les phrases comme critère premier de regroupement et la rapidité du processus. Le premier terme de ce compromis nous a conduit à écarter les algorithmes de partitionnement de type *k-means*, très dirigés par la donnée *a priori* d'un nombre de classes. Son second terme nous a fait mettre de côté les solutions qui, comme les algorithmes de clustering hiérarchique, nécessitent un temps de calcul très important pour calculer la matrice de similarité de toutes les phrases. Nous avons donc mis en œuvre un algorithme traitant les phrases une à une. Chaque nouvelle phrase est comparée à chacun des clusters déjà formés. Si la similarité entre cette phrase et le cluster le plus proche est jugée trop faible, un nouveau cluster est créé pour l'abriter ; sinon, la phrase est affectée au cluster qui lui est le plus proche. Afin de privilégier le critère d'équivalence sémantique, la similarité entre une phrase et un cluster est évaluée de façon conservative en retenant la similarité de cette phrase avec la phrase du cluster qui lui est la moins similaire. Pour éviter les effets de séquence, les phrases des documents sont présentées selon un ordre aléatoire.

Mesure de similarité sémantique : La problématique de la similarité sémantique de phrases a fait l'objet d'une attention particulière depuis quelques temps, matérialisée notamment par plusieurs tâches des dernières éditions de l'évaluation SemEval. Néanmoins, l'application de ces travaux au contexte du RA se heurte en pratique à un problème de ratio défavorable entre leur rapidité et la qualité de leurs résultats. À cet égard, nous avons testé deux options. L'une repose sur l'utilisation de plongements lexicaux (*word embeddings*), représentés en pratique par un ensemble de vecteurs lexicaux construits grâce à l'outil *word2vec* (Mikolov *et al.*, 2013) à partir de 100 millions de mots issus de Google News [1]. Chaque phrase est représentée par la moyenne des représentations des mots qui la composent, dimension par dimension, et la similarité de deux phrases est donnée par la mesure *cosinus* appliquée à leurs représentations. La seconde option est incarnée par le meilleur système de

1. Vecteurs GoogleNews-vectors-negative300.bin.gz : `https://code.google.com/archive/p/word2vec/`

l'évaluation SemEval 2014, fondé sur l'alignement des phrases au niveau lexical (Sultan *et al.*, 2014). Les expérimentations que nous avons menées montrent que la première approche est beaucoup plus rapide que la seconde, avec des performances comparables, voire supérieures. Les expérimentations de la section 4 se limitent donc à la première option.

3.2 Sélection des phrases

Comme modèle ILP initial, nous adoptons celui proposé par Gillick & Favre (2009). Son objectif est de maximiser la couverture en bigrammes du résumé final par rapport aux documents initiaux dans la limite d'une taille maximale du résumé. Le score d'un résumé est la somme des poids des bigrammes qu'il inclut, chaque bigramme y contribue par son poids une seule fois, ce qui assure implicitement une forme de non-redondance. Dans ce qui suit, ce système est noté *ilp*

$$\text{Maximise} : \sum_i w_i.c_i$$

$$\text{avec} : \sum_j s_j.l_j \leq L \quad (1)$$

$$s_j.Occ_{ij} \leq c_i, \forall i,j \quad (2)$$

$$\sum_j s_j.Occ_{ij} \geq c_i, \forall i \quad (3)$$

$$c_i \in \{0,1\}\ \forall i \text{ et } s_j \in \{0,1\}\ \forall j$$

Les équations ci-contre formalisent le problème ILP. La variable c_i indique la présence du concept i dans le résumé. w_i est le poids du concept i, égal au nombre de documents où le bigramme apparaît au moins une fois. La longueur de la phrase j est notée l_j et la longueur maximale du résumé est la constante L. La variable s_j indique la présence de la phrase j dans le résumé et Occ_{ij} indique l'occurrence du concept i dans la phrase j. La contrainte (1) garantit le non dépassement de la taille maximale du résumé tandis que les contraintes (2) et (3) en garantissent la cohérence.

Une première façon de prendre en compte le clustering sémantique est de sélectionner un représentant par cluster, par exemple la phrase la plus longue, supposée maximiser la couverture informationnelle, et d'appliquer le système *ilp*. Cette approche ne donne en pratique pas de bons résultats car elle réduit le choix du processus de sélection dont les critères d'optimisation, appliqués à une échelle globale aux bigrammes des textes, sont plus efficaces que cette stratégie locale.

Nous avons donc choisi d'intégrer la prise en compte du clustering sémantique directement au sein du problème ILP en y ajoutant une contrainte supplémentaire imposant de sélectionner au plus une phrase pour chaque cluster sémantique.

$$\sum_j s_j.Cls_{jk} \leq 1, \forall j,k \quad (4)$$

Cette contrainte est formalisée par l'équation (4) ci-contre au sein de laquelle la constante Cls_{jk} indique la présence de la phrase j dans le cluster k.

Cette approche est notée *ilp-sem-constr* dans ce qui suit. Cette façon de prendre en compte le regroupement sémantique des phrases ne spécifie rien en revanche sur la façon de sélectionner un représentant dans un regroupement ou même un regroupement par rapport à un autre en dehors des critères portant sur les bigrammes. Pour contrôler plus explicitement ce choix, nous avons modifié la fonction d'objectif globale afin d'y introduire une notion générique de représentativité des phrases. La nouvelle fonction d'objectif à maximiser devient :

$$\sum_i w_i.c_i + \sum_j rep_j.s_j \quad (5)$$

où rep_j fait référence à la représentativité de la phrase j. Les systèmes ayant cette fonction d'objectif sont notés *ilp-sem-obj* dans ce qui suit. Ils intègrent les mêmes quatre contraintes que *sem-ilp-constr*.

4 Expérimentations

4.1 Évaluation

Corpus d'évaluation Les méthodes considérées ici sont non supervisées mais comportent un certain nombre de paramètres que nous avons étalonnés sur un corpus de développement avant d'évaluer les méthodes sur un corpus de test. Nous avons choisi deux corpus de référence pour le RA multi-document en domaine générique issus des campagnes d'évaluation DUC : le corpus DUC 2003 pour l'entraînement et le corpus DUC 2004 pour le test. Ce dernier est composé de 50 clusters regroupant en moyenne 10 articles de presse relatifs au même sujet. La longueur moyenne de chaque article est de 570 mots.

Baselines Nous avons adopté comme référence basse trois systèmes de l'état de l'art en nous appuyant sur les résumés mis à disposition pour le corpus DUC 2004 par Hong *et al.* (2014). *freqsum* (Nenkova *et al.*, 2006) considère la fréquence des mots dans les documents à résumer comme indicateur de pertinence des phrases et gère la redondance en éliminant les phrases très similaires aux phrases sélectionnées pour le résumé. *centroid* (Radev *et al.*, 2004) représente chaque document par le barycentre des vecteurs *tf-idf* des phrases qui le constituent et favorise les phrases dont les vecteurs sont les plus proches de ce centroïde. *lexrank* (Erkan & Radev, 2004) construit un graphe dont les nœuds représentent les phrases des documents et les arcs, la similarité *cosinus* entre les phrases. Ces dernières sont ensuite classées suivant la pertinence obtenue en appliquant l'algorithme PageRank (Page *et al.*, 1999) sur le graphe construit. Enfin, le système *ilp* décrit à la section 3.2 nous sert de référence basse plus spécifiquement en lien avec le cadre de Gillick & Favre (2009) dans lequel nous nous situons. Dans ce cas, nous nous sommes appuyés sur une adaptation du système Potara (Bois, 2014) (`https://github.com/sildar/potara`).

Systèmes de l'état de l'art Nous avons également utilisé les résumés produits par 5 systèmes plus élaborés de l'état de l'art, toujours issus de Hong *et al.* (2014). Ces systèmes sont CLASYY11 (Conroy *et al.*, 2011), DPP (Kulesza & Taskar, 2012), OCCAMS_V (Davis *et al.*, 2012), RegSum (Hong & Nenkova, 2014) et Submodular (Lin & Bilmes, 2011).

Systèmes proposés Nous présentons 5 variantes de l'approche proposée à la section 3. Dans les 5 systèmes, les bigrammes formés par deux mots vides sont éliminés ainsi que les phrases de taille inférieure à 10 et les phrases entre guillemets. Les systèmes *ilp-sem-constr-df* et *ilp-sem-constr-sf* sont des variantes du modèle *ilp-sem-constr* n'intégrant que la contrainte supplémentaire (4). *ilp-sem-constr-sf* utilise comme poids des bigrammes le nombre de phrases dans lesquelles ils apparaissent tandis que *ilp-sem-constr-df* exprime ce poids en nombre de documents, comme toutes les autres variantes. Les trois autres systèmes sont des variantes du modèle *ilp-sem-obj* incluant la nouvelle fonction d'objectif (5). Dans *ilp-sem-obj-len-a*, nous attribuons une valeur de représentativité fixe élevée à la phrase la plus longue de chaque cluster et une valeur nulle aux autres phrases. Dans *ilp-sem-obj-len-b*, la représentativité des phrases d'un cluster est proportionnelle à leur longueur afin d'offrir plus de flexibilité à l'optimisation par ILP. Enfin, dans *ilp-sem-obj-pos*, la représentativité d'une phrase est déterminée en fonction de sa position dans le document à résumer. Des valeurs plus élevées sont attribuées aux phrases se situant au début ou à la fin des documents. Chacune des variantes présentées dépend par ailleurs du seuil de similarité minimale pour intégrer une phrase à un cluster, seuil fixé à 0,9 pour les quatre premiers systèmes et à 0,5 pour le cinquième. Dans ce dernier cas, l'influence du paramètre est moins sensible et la valeur choisie permet surtout, à performances comparables, de minimiser les temps de traitement. Le poids minimal des bigrammes est fixé à 1 dans

Méthode	F-mesure ROUGE			
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU4
Baselines				
freqsum	35,07	8,05	31,46	12,17
centroid	36,12	7,90	31,10	12,29
lexrank	35,94	7,48	31,38	11,93
ilp	37,41	8,55	29,18	12,72
Systèmes de l'état de l'art				
CLASYY11	37,32	9,22	32,74	13,17
DPP	39,76	9,60	30,83	**13,84**
OCCAMS_V	38,44	9,73	**34,48**	13,43
RegSum	38,39	9,70	34,10	13,78
Submodular	**39,25**	9,36	33,91	13,77
Systèmes proposés				
ilp-sem-constr-df	38,00	9,94	29,78	13,41
ilp-sem-constr-sf	37,22	9,59	30.63	13,11
ilp-sem-obj-len-a	38,30	**10,16**	30,22	13,63
ilp-sem-obj-len-b	38,28	**10,17**	30,30	13,63
ilp-sem-obj-pos	37,83	9,09	29,68	13,14

TABLE 1: F-mesures moyennes ROUGE des différents systèmes sur les données DUC 2004

les deux premiers systèmes et à 3 dans les autres.

4.2 Résultats et discussion

Nous présentons dans le tableau 1 les résultats de l'évaluation des différents systèmes considérés avec différentes variantes de la mesure ROUGE [2] qui mesure le recouvrement des n-grammes entre les résumés produits et des résumés de référence (Lin, 2004), les scores ROUGE-2 étant jugés les plus représentatifs d'après Hong *et al.* (2014). Les chiffres en gras indiquent le meilleur score pour chaque variante de ROUGE. Le test des rangs signés de Wilcoxon, recommandé par Rankel *et al.* (2011), a été effectué pour les scores ROUGE-2 et a montré avec un intervalle de confiance de 95% que tous les systèmes proposés, à l'exception de *ilp-sem-obj-pos*, ont des résultats supérieurs à toutes les baselines de façon statistiquement significative. L'amélioration significative par rapport au système *ilp* en particulier confirme l'intérêt du clustering sémantique dans l'élimination explicite de la redondance et le ciblage des choix du module ILP. À cet égard, le système ICSISumm (Gillick *et al.*, 2008) obtient des performances un peu supérieures à *ilp* avec les mêmes principaux généraux et pourrait de façon intéressante constituer une base permettant d'améliorer encore un peu nos résultats. Par ailleurs, nos deux meilleures configurations, *ilp-sem-obj-len-a* et *ilp-sem-obj-len-b*, ont des scores supérieurs aux systèmes de l'état de l'art présentés, même si ce dépassement n'est pas statistiquement significatif. De façon plus spécifique, ces résultats montrent aussi que dans notre cadre, le critère de longueur des phrases est plus pertinent que celui de position dans le document. Dans la mesure où les autres critères se focalisent sur la couverture informationnelle, il est vraisemblable que ce critère de représentativité d'une phrase soit plus perturbateur que bénéfique dans ce mode d'intégration.

En conclusion, nous avons réussi à montrer que l'intégration de la similarité entre phrases comme contrainte, dans un problème de maximisation de la couverture en bigrammes, apporte une amélioration significative. Il serait toutefois intéressant de considérer davantage de critères pour la sélection des phrases et en particulier d'associer les critères liés au contenu et ceux liés à la représentativité

2. Paramètres ROUGE : -n 2 -2 4 -m -l 100 -u -c 95 -p 0.5 -f A -r 1000

de ce contenu. Notre intégration du critère de position ne s'est à cet égard pas avérée positive mais n'avait qu'un caractère préliminaire. La conjugaison d'une méthode d'intégration plus élaborée et de critères plus profonds reposant sur la mise en évidence de la structure discursive des documents à résumer est une voie que nous envisageons pour progresser dans cette perspective.

Références

BOIS R. (2014). Multi-document Summarization Through Sentence Fusion. Master's thesis, Université de Nantes, France.

BOUDIN F., MOUGARD H. & FAVRE B. (2015). Concept-based Summarization using Integer Linear Programming: From Concept Pruning to Multiple Optimal Solutions. In *2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, p. 1914–1918, Lisbon, Portugal.

CONROY J. M., SCHLESINGER J. D. & KUBINA J. (2011). CLASSY 2011 at TAC: Guided and Multi-lingual Summaries and Evaluation Metrics. In *Fourth Text Analysis Conference (TAC 2011)*.

DAVIS S. T., CONROY J. M. & SCHLESINGER J. D. (2012). OCCAMS - An Optimal Combinatorial Covering Algorithm for Multi-document Summarization. In J. VREEKEN, C. LING, M. J. ZAKI, A. SIEBES, J. X. YU, B. GOETHALS, G. I. WEBB & X. WU, Eds., *ICDM Workshops*, p. 454–463: IEEE Computer Society.

ERKAN G. & RADEV D. R. (2004). LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research (JAIR)*, **22**, 457–479.

GILLICK D. & FAVRE B. (2009). A Scalable Global Model for Summarization. In *Workshop on Integer Linear Programming for Natural Langauge Processing (ILP'09)*, p. 10–18, Boulder, Colorado.

GILLICK D., FAVRE B. & HAKKANI-TÜR D. (2008). The ICSI Summarization System at TAC 2008. In *TAC 2008*.

HONG K., CONROY J., FAVRE B., KULESZA A., LIN H. & NENKOVA A. (2014). A Repository of State of the Art and Competitive Baseline Summaries for Generic News Summarization. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Eds., *Ninth International Conference on Language Resources and Evaluation (LREC'14)*, p. 1608–1616, Reykjavik, Iceland: ELRA.

HONG K. & NENKOVA A. (2014). Improving the Estimation of Word Importance for News Multi-Document Summarization. In *14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, p. 712–721, Gothenburg, Sweden.

KULESZA A. & TASKAR B. (2012). Determinantal Point Processes for Machine Learning. *Foundations and Trends in Machine Learning*, **5**(2–3).

LI C., LIU Y. & ZHAO L. (2015). Using External Resources and Joint Learning for Bigram Weighting in ILP-Based Multi-Document Summarization. In *NAACL-HLT 2015*, p. 778–787.

LI C., QIAN X. & LIU Y. (2013). Using Supervised Bigram-based ILP for Extractive Summarization. In *51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, p. 1004–1013, Sofia, Bulgaria.

LI P., WANG Y., GAO W. & JIANG J. (2011). Generating Aspect-oriented Multi-Document Summarization with Event-aspect model. In *2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, p. 1137–1146, Edinburgh, Scotland, UK.

LIN C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In *ACL-04 Workshop Text Summarization Branches Out*, p. 74–81, Barcelona, Spain.

LIN H. & BILMES J. A. (2011). A Class of Submodular Functions for Document Summarization. In D. LIN, Y. MATSUMOTO & R. MIHALCEA, Eds., *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011)*, p. 510–520.

MCDONALD R. (2007). A Study of Global Inference Algorithms in Multi-document Summarization. In *29th European Conference on IR Research (ECIR'07)*, p. 557–564: Springer-Verlag.

MIKOLOV T., YIH W.-T. & ZWEIG G. (2013). Linguistic Regularities in Continuous Space Word Representations. In *2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2013)*, p. 746–751, Atlanta, Georgia.

NENKOVA A., VANDERWENDE L. & MCKEOWN K. (2006). A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization. In E. N. EFTHIMIADIS, S. T. DUMAIS, D. HAWKING & K. JÄRVELIN, Eds., *29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2006)*, p. 573–580: ACM.

PAGE L., BRIN S., MOTWANI R. & WINOGRAD T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66, Stanford InfoLab.

RADEV D., ALLISON T., BLAIR-GOLDENSOHN S., BLITZER J., ÇELEBI A., DIMITROV S., DRABEK E., HAKIM A., LAM W., LIU D., OTTERBACHER J., QI H., SAGGION H., TEUFEL S., TOPPER M., WINKEL A. & ZHANG Z. (2004). MEAD — A platform for multidocument multilingual text summarization. In *3rd Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.

RANKEL P., CONROY J. M., SLUD E. V. & O'LEARY D. P. (2011). Ranking Human and Machine Summarization Systems. In *2011 Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*, p. 467–473.

SULTAN M. A., BETHARD S. & SUMNER T. (2014). DLS@CU: Sentence Similarity from Word Alignment. In *8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 241–246, Dublin, Ireland.

WOODSEND K. & LAPATA M. (2012). Multiple Aspect Summarization Using Integer Linear Programming. In *2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*, p. 233–243, Jeju Island, Korea.