UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MÉTHODES HYBRIDES POUR LE RÉSUMÉ AUTOMATIQUE DE TEXTE PAR EXTRACTION ET ABSTRACTION

THÈSE
PRÉSENTÉE COMME EXIGENCE PARTIELLE
DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR

ALAIDINE BEN AYED

MAI 2022

REMERCIEMENTS

Je tiens à remercier tout d'abord mes directeurs de thèse Ismaïl Biskri et Jean-Guy Meunier pour leur disponibilité, leurs conseils et leur implication dans cette thèse.

Mes remerciements vont également à Juan-Manuel Torres- Moreno, Serge Robet et Daniel Lemire pour avoir accepté de rapporter sur cette thèse et à Louise Brissette pour sa participation au jury.

Je remercie vivement aussi les chercheurs des laboratoires LANCI (Laboratoire d'analyse cognitive de l'information) et LI2A (Laboratoire d'Intelligence Artificielle appliquée) pour leurs conseils précieux lors de la soutenance à mi-parcours.

Enfin, je tiens à exprimer ma gratitude à ma famille notamment mes parents, mes frères et ma fille.

DÉDICACE

À mon père El-Kamel, ma mère Samira, mes deux frères Atef et Ayoub, ma sœur Marion, et à ma chère fille Lyn

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

RÉSUMÉ

La présente recherche s'inscrit dans le cadre du traitement automatique du langage naturel. Dans ce travail, nous nous intéresserons aux résumés automatiques de textes en portant une attention particulière à la famille de méthodes de résumé automatique par compréhension. L'objectif principal de cette recherche était de concevoir et d'implémenter deux approches de résumé automatique par extraction et par abstraction ainsi qu'un protocole d'évaluation de résumés automatiques de texte par compréhension.

Le volet informatique de notre recherche présente de manière détaillée les approches proposées pour le résumé automatique de texte. En ce qui concerne le volet cognitif, nous accordons une attention particulière aux théories de compréhension de textes issues de la psychologie cognitive. L'hypothèse générale qui sous-tend notre recherche est qu'un modèle informatique de résumé automatique de textes ou d'évaluation de résumé automatique doit être inspiré de l'activité résumante chez le sujet humain.

Mots-clés : Résumés automatiques de textes par compréhension, Résumés automatiques par extraction et abstraction, Méthodes d'évaluation des résumés automatiques. Modèles issus de la psychologie cognitive pour la compréhension de textes.

ABSTRACT

This work is part of the NLP (Natural Language Processing) research. We will pay particular attention to the by-comprehension text summarization approaches. This research aims to design by-comprehension extractive and abstractive models and an evaluation protocol of automatic text summarization.

The computer science component of our research resides in our proposed approaches for automatic text summarization details. As for the cognitive component, special attention is given to the cognitive psychology theories for text summarization. The general hypothesis underlying our research is that an automated model of text summarization, also an evaluation protocol for the same purpose, should be inspired by the human subject summarizing activity.

Keywords: Automatic text Summarization by comprehension, By extraction/ abstraction text summarization, Evaluation methods of automatically generated summaries. Cognitive psychology models of text comprehension

CHAPITRE I : INTRODUCTION GÉNÉRALE

1.1 Le contexte

L'idée de l'automatisation de l'activité résumante est née en raison des économies qu'elle pourrait engendrer. En effet, le volume de données textuelles générées (textes techniques et scientifiques, articles de presse, etc.) fait de l'analyse de ces derniers une tâche ardue. Afin de mieux cerner les défis à relever lors de la conception et de la réalisation d'une nouvelle approche de résumé automatique de textes, nous devons tout d'abord nous attarder aux caractéristiques fondamentales de ces textes dans une optique de Big Data (données massives).

1.1.1 Le volume des données générées

Les statistiques révèlent que sur les 7,6 milliards d'habitants de notre planète, 4,1 milliards sont des internautes (environ 40 % de la population mondiale) (Cox, 2018). Le web ainsi que la numérisation de documents finissent par générer des données massives qui présentent de nouveaux défis d'analyse et de traitement (Smith, 2018). Il faut rappeler que dans le cadre de ce projet de recherche, nous nous intéressons à un défi spécifique, à savoir les méthodes de résumé automatique de textes.

1.1.2 La vitesse

La vitesse à laquelle d'énormes quantités de données textuelles sont générées par les échanges sur les médias sociaux et la numérisation pose un grand défi en ce qui concerne les méthodes de résumé automatique de textes. En revanche, les outils de TAL doivent prendre en considération de multiples dimensions, y compris les liens spatiaux et temporels ainsi que la non-hétérogénéité des données qui viennent de multiples sources, ce qui fait de l'automatisation de la création des résumés une tâche ardue.

1.1.3 La variété

Souvent, ce problème émerge quand il s'agit de textes extraits des médias sociaux où les utilisateurs n'emploient pas la même langue ni le même style. Ainsi, ces textes n'obéissent pas aux règles d'orthographe et de grammaire classiques. Dans le cadre de cette recherche, le problème de variété ne se pose pas, car nous travaillons sur des articles de journaux qui obéissent aux règles de grammaire et d'orthographe classiques.

1.2 Les enjeux d'un résumé automatique par compréhension

Un bon résumé doit être doté d'un certain nombre de caractéristiques, à savoir la couverture, la fidélité et la cohérence :

- La couverture : Il s'agit d'un critère important pour caractériser un bon résumé. Elle présente en quelque sorte le rapport entre les propos énoncés dans le texte source et ceux présents dans le résumé généré.
- La fidélité : Elle peut être considérée comme une mesure de la qualité globale du résumé. Elle dépend systématiquement de la couverture. En effet, si la couverture est respectée, le résumé généré sera assez fidèle au texte source.
- La cohérence : Elle peut être obtenue à la suite de l'application des techniques de suivi du thème, qui font en sorte que le résumé généré soit cohérent.

Notez que « résumer » revient à identifier, puis extraire l'information la plus saillante d'un texte source dans le but de le transformer en une version abrégée. La taille du résumé est inférieure à celle du texte source. En effet, celui-ci doit contenir les idées et énoncés présents dans le texte d'origine (Mani, 1999). Donc, le terme « activité résumante » désigne une démarche cognitive complexe qui fait appel à différents procédés de réduction afin de retenir l'essentiel du texte à résumer. Ainsi, le sujet humain fait intervenir plusieurs capacités intellectuelles afin de dissocier ce qui est facultatif de ce qui est essentiel. Dans ce qui suit, nous exposons le processus entrepris par un humain pour générer un résumé par compréhension.

Selon (Kumar, 2016), l'élaboration du résumé par compréhension passe par cinq phases : *1)* la lecture du texte source, *2)* l'analyse de l'information, *3)* la hiérarchisation de l'information, *4)* la synthèse de l'information et *5)* la rédaction du résumé. Nous relatons en détail chacune de ces phases dans ce qui suit.

La phase de lecture complète du texte source : Durant cette phase, le sujet humain effectue une ou plusieurs lectures du texte source. Il s'assure de la compréhension parfaite de son contenu et étudie le vocabulaire employé. Souvent, il met l'accent sur certaines parties, c'est-à-dire le début et la fin.

L'analyse de l'information source : Une fois que la lecture du texte source est achevée, on passe à l'analyse de l'information, soit l'analyse de certains segments du texte, des mots-clés, des connecteurs, etc. Le lecteur effectue le bilan de toutes les informations contenues dans le texte source. Il dégage l'ensemble des propos énoncés dans le texte tout en restant neutre et objectif. Il peut ensuite marquer les segments les plus informatifs qui constitueront par la suite un simple résumé par extraction ou la base d'un résumé par abstraction plus élaboré.

La hiérarchisation de l'information : Cette phase est en quelque sorte une opération de tri des différentes informations énoncées dans le texte. En effet, l'humain classe les notions fondamentales pour la bonne compréhension du texte et s'affaire à dégager les liaisons sémantiques exprimant la cohérence entre elles.

La synthèse de l'information : Durant cette phase, le sujet humain réalise un filtrage de l'information recueillie lors de la phase précédente. Ce filtrage peut s'opérer à l'aide des mécanismes cognitifs de suppression, d'intégration, de généralisation et/ou par l'application de procédés stylistiques et linguistiques permettant le filtrage de mots et la compression des phrases.

La rédaction du texte du résumé : Il s'agit de projeter les idées dégagées du texte original dans une structure narrative, tout en restant neutre et en veillant à la clarté des idées. Le sujet humain peut reprendre les mêmes phrases employées dans le texte original tout comme il peut employer ses propres mots. Dans tous les cas, le résumé généré doit satisfaire les trois critères d'un bon condensé textuel, à savoir la couverture, la fidélité et la cohérence.

1.3 Problématique

Nous démontrerons dans le chapitre portant sur l'état de l'art que la plupart des modèles proposés dans le cadre de la recherche en méthodes de résumé automatique de textes sont issus d'approches classiques purement statistiques, linguistiques, ou d'une combinaison des deux. Dans le cadre de cette recherche, nous nous basons sur les travaux de compréhension de textes réalisés en psychologie cognitive pour concevoir et implanter de nouveaux protocoles de résumés automatiques et d'évaluation de résumés automatiques de textes. Deux défis majeurs sont à relever lors de l'élaboration de la version computationnelle pour un modèle théorique de compréhension de texte :

i. La simulation des processus cognitifs ; c'est-à-dire les *macro-règles* qui interviennent pour assurer la bonne compréhension du texte lu.
ii. La simulation d'un système de mémoire complexe qui véhicule les différents concepts lors de la lecture du texte à résumer.

En ce qui concerne la simulation des processus cognitifs, il est parfois difficile d'automatiser certaines tâches. À titre d'exemple, la conversion du texte brut en un ensemble d'unités sémantiques appelées propositions peut induire des erreurs qui auront pour effet de rendre les traitements ultérieurs plus difficiles. Quant à la simulation du système de mémoire, notons que celle-ci est organisée en plusieurs niveaux :

La mémoire à court terme : Souvent appelée la mémoire du présent. Nous l'utilisons pour retenir des informations dans le cerveau durant de très courtes durées.

La mémoire de travail (mémoire immédiate) : Elle correspond à une sorte de mémoire cache souvent utilisée à court terme. Elle joue un rôle essentiel lorsque nous voulons effectuer deux choses en parallèle, comme parler à quelqu'un en lisant un texte.

La mémoire à long terme : C'est une mémoire permanente où les informations sont stockées pendant une longue durée qui peut aller jusqu'à la durée de vie de l'être humain. Dotée d'une capacité abyssale, elle se décompose en deux sous-systèmes de mémoire :

- La mémoire explicite ou mémoire déclarative, qui mémorise les informations que nous pouvons exprimer par le biais du langage ; un souvenir personnel par exemple. On distingue deux types de mémoire explicite, soit la mémoire sémantique et la mémoire épisodique. La première sert d'entrepôt des connaissances générales sur soi (souvenirs) et notre environnement (géographie, nature, et même les noms des objets, leurs fonctions, leurs utilisations ou leurs caractéristiques). En revanche, la deuxième stocke les informations liées aux événements vécus et leur contexte (le lieu, la date ou l'état émotionnel).

- La mémoire implicite, aussi appelée mémoire non déclarative ou mémoire procédurale, qui permet l'acquisition et l'utilisation de compétences motrices (par exemple faire du vélo ou pratiquer un sport).

La simulation de ce système de mémoire complexe semble constituer l'un des défis majeurs avec lequel il faut composer lors de l'implantation de la version computationnelle d'un modèle théorique de compréhension de textes puisqu'il sert de support principal pour l'information véhiculée dans ce système

1.4 Volets informatiques et cognitifs de la thèse

Le volet informatique de notre recherche présente de manière détaillée les approches proposées pour le résumé automatique de texte. En ce qui concerne le volet cognitif, nous accordons une attention particulière aux théories de compréhension de textes issues de la psychologie cognitive. Nous prétendons qu'un modèle informatique de résumé automatique de textes ou d'évaluation de résumé automatique doit être inspiré de l'activité résumante chez le sujet humain.

1.5 Hypothèses

L'hypothèse principale qui sous-tend notre recherche est qu'on peut améliorer la qualité d'un résumé en simulant par des algorithmes spécifiques certaines des opérations cognitives de compréhension de texte. Cette hypothèse se déploie en deux sous-hypothèses.

Hypothèse cognitive (H1) : Le résumé par extraction ou abstraction ou l'évaluation d'un résumé automatique est une opération cognitive liée à des processus de compréhension de textes ou de discours.

Hypothèse informatique (H2) : On peut simuler algorithmiquement l'opération cognitive de compréhension de texte afin d'élaborer des méthodes d'évaluation et des résumés par extraction et abstraction pertinents qui satisferont les trois contraintes de couverture, fidélité et cohérence ainsi que des modèles d'évaluation de résumés automatiques de textes pertinents.

1.6 Plan de la thèse

Ce travail est organisé de la façon suivante : Tout d'abord, nous décrivons l'état de l'art du domaine dans le prochain chapitre. Nous nous focalisons surtout sur les méthodes extractives et abstractives du résumé automatique ainsi que les méthodes d'évaluation de résumés automatiques de textes. Ensuite, nous présentons deux protocoles de résumé automatique par extraction et par abstraction dans les troisième et quatrième chapitres. Les deux approches reposent sur un modèle de compréhension de textes issu des travaux réalisés dans le cadre de la psychologie cognitive. Par la suite, nous proposons une approche cognitive d'évaluation de résumés automatiques par extraction dans le cinquième chapitre. Nous consacrons le sixième chapitre aux conclusions tirées de ce travail ainsi qu'aux directives et directions de recherche en traitement automatique du langage naturel dans le cadre de l'informatique cognitive. Dans l'annexe, nous décrivons un cas d'utilisation des résumés automatique comme outil d'analyse des documents pour implémenter un moteur de recherche basé sur l'outil LUCENE.

CHAPITRE II : ÉTAT DE L’ART

2.1 Introduction

Un résumé automatique de texte est un condensé d'un texte source plus long. Ce condensé est obtenu avec perte d'information au moyen de techniques informatiques basées sur des approches statistiques, linguistiques ou combinaison des deux. Il est à noter que produire un résumé humainement crédible demande au résumeur de soigneusement sélectionner et assembler des segments d'information selon leurs degrés de saillance. En revanche, ceci demande aussi de résoudre les problèmes de redondance, cohérence et cohésion afin de produire des condensés automatiques pertinents et de qualité.

Dans ce chapitre, nous présentons les différents types du résumé automatique. Ensuite nous faisons l'état de l'art des techniques d'évaluation des résumés automatiques.

2.2 Taxonomie des méthodes de résumé automatiques

Plusieurs ébauches en résumé automatique de texte ont été proposées depuis les années cinquante. Ci-dessous, nous effectuons une taxonomie de ces travaux selon la nature de l'entrée, la nature de la sortie, ainsi que la stratégie de construction du résumé automatique.

2.2.1 Selon le document d'entrée

Selon La forme : Le type, la structure, la granularité et le médium

La forme contient plusieurs aspects : le type, la structure, la granularité et le médium. La technique à employer pour générer le résumé automatique dépend souvent du type du document source qui peut être un article de journal (Sethi, 2017, Sahni, 2018 et Ben Ayed, 2019), un roman (Kazantseva, 2006, 2010), un rapport scientifique (Paice,1993, Zhang, 2019, Ramu, 2019), une pièce d'opinion (en Anglais : reviews) (Zhan et al, 2009, Mirani, 2017, Lovinger, 2019), un microblog ou des tweets (Dutta, 2019 et Chakraborty, 2019).

La structure comprend deux niveaux : l'organisation en sections du document à résumer ainsi que l'organisation et les relations rhétoriques entre les phrases et les mots employés. Des textes particuliers comme les articles scientifiques ont souvent une structure commune : une introduction des sections pour les travaux reliés, le travail présenté, l'évaluation, et la conclusion finale. Ces sections sont généralement marquées par des en-têtes (Paice, 1993 et Jaidka, 2013). Souvent, on a intérêt à ce que le résumé généré ait la même structure que celle du texte source, comme dans le travail de Farzindar (Farzindar, 2004), qui a généré des résumés automatiques des textes juridiques en se basant sur leur structure qui comporte quatre sections : une introduction, une section contexte, une section raisonnement juridique et une conclusion. Il est à noter aussi que la rhétorique du texte à résumer peut-être modélisée par l'intermédiaire de différentes théories de l'organisation du texte, comme la technique de la structure rhétorique (RST : *Rhetorical structure theory*). La RST définit les relations entre les phrases du texte (exp. une déclaration suivie d'une élaboration). Nous l'avons utilisée dans (Ben Ayed, 2019) pour assurer la cohérence des résumés générés.

En résumé automatique de textes, on peut travailler sur des récits courts comme on peut travailler sur des romans. La transformation du contenu dépend toujours de la granularité qu'on veut atteindre qui dépend du degré de réduction souhaité. Les auteurs dans (Woodsend, 2010 et Kazantseva, 2006) ont proposé une approche de résumé automatique d'histoires qui a pour but de donner un condensé compressé offrant un aperçu sur le thème général sans devancer trop de détail concernant les caractères, les lieux, etc.

Souvent, le texte est le support le plus utilisé en résumé automatique, bien qu'il existe d'autres supports comme l'image (Carson, 2002, Fei-Fei, 2003 et Trieu, 2020), l'audio (Hori, 2001, Inoue, 2004 et Gonz'alez-Gallardo, 2020) et la vidéo (Ekin, 2003, Albanese, 2006 et Wei, 2018). Le résumé automatique d'image consiste souvent à extraire des contours, des régions ou des formes de l'image à traiter, initialement présentée sous forme de pixels. Le résumé automatique de documents audio requiert souvent une phase de prétraitement qui consiste à transcrire des documents audio sources. Ensuite, ces documents sont traités comme des textes classiques. Enfin, le résumé automatique de vidéos est une combinaison des deux premiers.

Selon l'unité : Résumés automatiques mono-documents vs résumés automatiques multi-documents

En se basant sur le nombre de documents à traiter, on distingue deux catégories d'approches de résumé automatique : mono-document et multi-documents. Un résumé mono-document est un résumé généré à partir d'un seul document, même si ce dernier est produit en compilant plusieurs documents. Le résumé automatique mono-document a fait l'objet de plusieurs recherches, à commencer par le travail de Luhn (Luhn, 1958). Parmi ces recherches, on peut citer : (Edmundson, 1969, Mitkov, 1993, Kupiec, 1995, Hovy, 1998, Farzindar, 2004, Joshi, 2018 et Ben Ayed, 2019), etc. À l'opposé, un résumé multi-documents est généré à la suite de l'analyse du contenu de plusieurs documents en entrée, ayant un ou plusieurs thèmes en commun (McKeown, 1995, Boudin, 2009, Gupta, 2012, Sahoo, 2016 et Tomer, 2021). Le résumé automatique multi-documents présente d'avantage de défis tels que la gestion de la redondance, des relations temporelles, etc. (Gupta, 2012).

2.2.2 Selon le document de sortie

Selon la partialité : résumé automatique neutre vs résumé automatique évaluatif

Un résumé automatique selon la partialité peut être neutre ou évaluatif. Généralement, ce critère est pris en considération quand il s'agit de textes d'opinion. Un résumé neutre est un condensé du (des) contenu (s) de document(s) d'entrée, généré sans y introduire des critiques ou des jugements supplémentaires. En revanche, un résumé évaluatif introduit des déclarations d'opinion. À titre d'exemple, les auteurs dans (Généreux, 2009, Dutta, 2019 and Chakraborty, 2019) ont proposé une approche de résumé automatique de blogs, qui crée un condensé d'opinions reportées dans des blogs tout en prenant une position sur les faits considérés.

Selon le public : Résumé automatique générique vs résumé automatique orienté usager

Selon le public, un résumé automatique peut être générique ou orienté par requête utilisateur. Un résumé générique est un résumé qui synthétise un ou plusieurs documents d'entrée, en donnant le même ordre d'importance pour tous les thèmes cités, et sans tenir compte de ce que l'utilisateur cherche (Kazantseva, 2006). À l'inverse, un résumé orienté usager est un résumé qui doit mettre l'accent sur le (les) thème (s) recherchés par l'utilisateur. Ce type de résumé demande une analyse en profondeur et une compréhension des thèmes figurant dans le texte à résumer et dépend fortement des contextes, de l'audience et des cadres spatio-temporels (Spärck Jones, 2007). (Rahman, 2020) a proposé plusieurs techniques de désambiguïsation pour gérer les cas d'ambiguïtés en lien avec le contexte et les cadres spatio-temporels afin d'améliorer la qualité des résumés orientés usager. Les travaux de (Reeve, 2007) qui s'intéresse aux textes biomédicaux et (Farzindar, 2004) qui prend en entrée des textes juridiques sont des exemples de travaux portant sur le résumé automatique orienté usager.

Selon la fonction : Résumé automatique indicatif vs résumé automatique informatif

Un résumé est soit informatif, soit indicatif. Le résumé informatif est un condensé du texte original reportant le plus largement possible les thèmes du document source. Les approches proposées par (Boudin, 2009 et Usunier, 2005) sont respectivement des exemples de méthodes de résumés automatiques informatifs multi-documents et informatifs mono-documents. À l'inverse, un résumé indicatif évoque les thèmes les plus saillants évoqués par le texte à résumer. Il faut souligner que la création d'un résumé automatique indicatif peut se restreindre à l'identification des mots-clés et thèmes présents dans le document source, l'approche proposée par (Liu, 2009, Li, 2019) en étant un exemple.

La combinaison des deux techniques de résumés automatiques indicatifs et informatifs peut servir d'outil pour créer des systèmes de résumés guidés. Par exemple, les auteurs dans (Saggion, 2002) ont proposé un système de résumé automatique qui génère initialement un résumé indicatif. Ensuite, l'utilisateur choisit parmi les thèmes proposés celui qui l'intéresse. Un résumé informatif est généré ensuite en se basant sur la requête de l'utilisateur. La requête dans ce contexte est l'ensemble des sujets sélectionnés à partir du résumé indicatif.

2.2.3 Selon la stratégie de construction

Résumé automatique extractif vs résumé automatique abstractif

Un résumé par extraction évite la génération de texte. Ceci permet d'une part, de se concentrer sur la sélection du contenu pertinent et d'autre part, d'obtenir un résumé lisible et linguistiquement correct (Abualigah, 2020). La cohérence n'est toutefois pas garantie. Par exemple, si le système de résumé sélectionne des phrases contenant des références (acronyme, pronom personnel, etc.) et ne sélectionne pas les phrases contenant leurs antécédents, il est fort probable que le résumé produit soit incompréhensible. Pour pallier ce problème, certains travaux considèrent le paragraphe comme unité d'extraction au lieu de la phrase (Salton, 1996). Ceci permet de garder la cohérence du texte source, mais ne peut pas être applicable dans le cas de résumés courts. De plus, il est évident que cette méthode réduit la précision du résumé en y incluant des phrases peu importantes juste pour améliorer la cohérence. D'autres chercheurs procèdent à des étapes de pré/post-traitement du texte qui améliorent partiellement la cohérence globale du résumé, comme la résolution des références anaphoriques dans le texte source (Trandabâț, 2011). Le processus principal dans le résumé extractif est la sélection des segments de textes (généralement les phrases) pertinents et non redondants sans dépasser une taille limite du résumé. Ce principe limite la couverture des informations apportées par le texte source (Ben Ayed, 2019). Les résumés abstractifs souffrent moins de ce problème puisque l'information peut y être reformulée.

Les méthodes de résumé abstractives imitent, jusqu'à un certain degré, le processus naturel accompli par les humains pour résumer un document. Par conséquent, elles produisent des résumés plus similaires aux résumés manuels. Ce processus peut être décrit par deux étapes majeures : la compréhension du texte source et la génération du résumé (Khan, 2014). Ces deux tâches sont assez complexes. C'est pourquoi elles ont été simplifiées. La première étape vise à analyser sémantiquement le contenu du texte et à identifier les parties à exprimer dans le résumé. Elle a parfois pris la forme d'une tâche d'extraction d'information liée au domaine abordé (Genest, 2011, 2012) ou de regroupement des phrases du texte source (Filippova, 2010). La génération de texte est un domaine en soi. Une des approches simplifiées consiste à appliquer des techniques de génération text-to-text : utilisation de paraphrases (Madnani, 2010) ou fusion et compression de phrases (Filippova, 2010). Une alternative consiste à induire un modèle textuel du domaine (patron) et de l'instancier lors de la génération (Cheung, 2013). Les nouvelles techniques de résumé automatiquement par abstraction sont principalement basées sur l'apprentissage profond (See, 2017, Paulus, 2017 and Chen, 2021).

Approches statistiques : Le travail de Luhn (Luhn, 1958) représente la première ébauche de résumé automatique. Il s'agit de la première méthode statistique basée sur la fréquence de mots. Les méthodes statistiques procèdent souvent par le calcul d'un score associé à chaque unité (phrase) afin d'estimer son importance sans tenir compte des aspects linguistiques. Le calcul d'un poids de saillance pour chaque phrase du texte source se base sur : *(a)* le calcul de similarité entre les phrases, et *(b)* un tas d'hypothèses Adhoc.

L'approche de résumé mono-document TextRank (Mihalcea, 2004) est l'approche de référence des méthodes basées sur le calcul de similarité entre les phrases. Cette approche utilise souvent la *tf-idf*, une fonction de la fréquence de mots dans le texte. Le score attribué à chaque unité textuelle (phrase) est la somme des scores des mots significatifs contenus dans celle-ci. L'hypothèse fondamentale sur laquelle reposent ces approches stipule que les phrases qui contiennent les mots les plus fréquents sont censées encoder la thématique du texte. Selon (Minel, 2003) cette hypothèse n'est pas toujours vraie quand il s'agit d'articles de presse où l'auteur a tendance à utiliser plusieurs termes lexicaux pour désigner un même référent en employant des synonymes ou de la métaphore, etc. En outre, l'utilisation des outils d'étiquetage morpho-syntaxique s'avère indispensable afin de résoudre les cas d'ambiguïté, ce qui rend ces systèmes dépendants des langues. De plus, l'extrait généré ne contient, par construction, que les phrases encodant la thématique principale du texte. Ainsi, le principe de la sélection d'unités textuelles par calcul de scores est utile quand il s'agit de résumés indicatifs, contrairement aux résumés informatifs. Afin de générer des condensés cohérents, le paragraphe peut être choisi comme unité textuelle à la place de la phrase puisqu'il offre plus de contexte (Salton, 1997). Ensuite, on calcule la similarité entre les paragraphes. Un lien de similarité entre deux paragraphes P_1 et P_2 est établi si le coefficient de similarité $Sim\ (P_1,\ P_2)$ est inférieur à un seuil donné. Plusieurs stratégies de sélection des paragraphes sont possibles par la suite, telles que la stratégie du « *busy path* » qui consiste à rechercher le paragraphe qui possède le plus de liens de similarité avec les autres, donc ceux qui traitent les principaux thèmes. Ce processus de recherche de paragraphes ayant le plus de liens de similarité se fait d'une façon itérative jusqu'à attendre le taux de compression souhaité.

Dans le contexte de résumé automatique multi-documents, les documents sont généralement représentés par des matrices dont les lignes sont des vecteurs pondérés avec la mesure *tf-idf* des mots significatifs qu'elles contiennent. Les documents sont regroupés selon la similarité de leurs matrices (Sammut, 2010). Les phrases les plus similaires au barycentre du cluster sont considérées comme représentatives du groupe de documents (Radev, 2004 et Neto, 2003), ce qui constitue un critère de sélection important. De nombreuses mesures de similarité textuelle utilisées en résumé automatique de textes ont été proposées dans le contexte classique de sélection de phrases pertinentes, ainsi que dans d'autres contextes plus complexes comme l'élimination de la redondance (Bär, 2015).

La sélection d'unités textuelles saillantes peut également se faire en se basant sur un tas d'hypothèses Adhoc : par exemple, les phrases qui contiennent des mots qui figurent dans le titre, sont généralement liées au thème principal étant donné que dans la majorité des cas, le titre informe de façon très brève sur le contenu principal du texte (Edmundson,1969). La similarité avec les sous-titres est aussi considérée comme indicateur de pertinence. De plus, les phrases se trouvant au début du texte source donnent fréquemment un aperçu général du contenu du texte. En outre, les phrases situées au début de chaque paragraphe sont souvent plus informatives que les autres (Lin, 1997 et McKeown, 1999). Les phrases très courtes sont aussi généralement peu informatives contrairement aux phrases très longues qui posent à leur tour des problèmes de redondance. Cette caractéristique qui favorise les phrases composées de 15 à 30 mots a été combinée avec d'autres pour produire des résumés automatiques par extraction (Schiffman, 2002).

Approches linguistiques : L'approche linguistique emploie plusieurs théories telles que la théorie de la structure rhétorique, et les chaînes lexicales. Ces modèles créent initialement une représentation de l'entrée. Le processus de création du résumé automatique consiste à réduire cette représentation en ayant recours à plusieurs règles de réduction, soit en gardant les segments les plus informatifs ou en créant une nouvelle représentation. Un résumé par extraction est souvent créé en fusionnant les phrases extraites. Finalement, un résumé par abstraction est obtenu en transformant la représentation réduite à un condensé par abstraction.

La méthode des chaînes lexicales a été introduite par (Barzilay, 1997). Une chaîne lexicale (*lexical chain* en anglais) est une séquence de mots ayant une distance courte (les mots qui apparaissent dans la même phrase ou dans des phrases adjacentes), ou une distance longue (le texte entier). Le principe de cette approche repose sur l'identification des chaînes lexicales puissantes qui sont censées apparaître dans les segments les plus informatifs. L'avantage de cette méthode réside dans le fait qu'elle permet de modéliser la cohésion lexicale du texte source en identifiant les mots qui sont sémantiquement reliés (synonymes, antonymes) même s'ils ne sont pas employés dans la même phrase. Elle permet également d'identifier le sens des phrases ainsi que les différents concepts et les liens entre eux. Un thésaurus (*Wordnet*) est utilisé pour l'identification des chaînes lexicales (Miller, 1995).

Les approches linguistiques peuvent également employer les théories de la structure rhétorique comme la RST (*Rhetorical Structure Theory*). Cette technique permet d'identifier deux types de phrases les noyaux et les satellites. Les noyaux sont les phrases principales. Les satellites sont les segments qui sont souvent employés pour détailler les propos exprimés dans les phrases noyaux. La RST permet également d'employer les relations rhétoriques (élaboration, justification, exemplification, etc.). (Marcu, 1998) a proposé d'extraire un sous-ensemble des phrases noyaux qui satisfait à un tas de critères classiques précédemment proposés par les approches statistiques. Contrairement à l'approche précédente qui étudiait les relations rhétoriques entre phrases de textes, (Ono, 1994) a proposé une méthode de résumé automatique basée sur la compression des phrases. Les relations entre les parties de la phrase sont modélisées par l'intermédiaire d'un arbre binaire. Ces structures binaires sont taillées pour ne garder que les parties les plus importantes de la phrase en se basant sur les relations rhétoriques entre les syntagmes nominaux et verbaux qui la composent.

2.3 Les techniques d'évaluation de résumé automatique

L'évaluation des résumés automatiques est un domaine de recherche actif. Plusieurs campagnes d'évaluation ont été organisées au fil des années. Les principales sont DUC, TAC et TREC. Au début, ces compagnes reposaient sur l'évaluation, par les lecteurs, de la qualité linguistique et du contenu du résumé généré. Cette évaluation est soit objective (en estimant la similarité des résumés candidats avec un résumé manuel) ou subjective (en jugeant la qualité du résumé sans se référer à un modèle). Ces techniques d'évaluation nécessitent une intervention de l'humain. Ainsi, plusieurs recherches ont proposé des métriques d'évaluation semi-automatique afin de faciliter la comparaison des différentes approches de résumé automatique. Ces métriques ne s'intéressent pas à la qualité linguistique ou grammaticale des résumés générés. Elles évaluent principalement le contenu sélectionné en le comparant à un modèle de référence. Ces techniques semi-automatiques stipulent donc la disponibilité des résumés de référence faits par des humains.

Dans ce qui suit, nous présentons : ROUGE, PYRAMID, et BE. ROUGE est la méthode d'évaluation semi-automatique la plus standard pour le cas du résumé automatique par extraction. Ensuite, nous faisons l'état de l'art des travaux portant sur les techniques proposées pour l'évaluation des résumés automatiques par abstraction.

2.3.1 ROUGE

ROUGE évalue la qualité d'un résumé généré par extraction en le comparant à un (ou plusieurs) résumé(s) de référence. Elle utilise principalement trois métriques pour quantifier les recouvrements entre les N-grammes deux textes. Ces trois métriques sont :

- La précision : C'est le rapport d'N-grammes communes au résumé généré automatiquement et aux résumés de référence sur la cardinalité (le nombre de N-grammes) du résumé automatique.
- Rappel. C'est le rapport d'N-grammes communes au résumé généré automatiquement et aux résumés de référence sur la cardinalité (le nombre de N-grammes) du résumé de référence.
- F-mesure (ou encore F-score): Il s'agit de la moyenne harmonique de la précision et du rappel.

Jusqu'à présent ROUGE est la mesure d'évaluation la plus utilisée en résumé automatique par extraction étant donné qu'elle a montré une forte corrélation avec les jugements humains (Lin, 2004). Notez aussi que d'autres variantes de ROUGE ont été proposées. Ces dernières exploitent la plus longue sous-séquence commune ou les bi-grammes distants et la dernière implémentation de ROUGE permet de calculer en plus la précision et la f-mesure (Hong, 2014).

2.3.2 PYRAMID

Le principe de la méthode *PYRAMID* repose sur l'annotation des résumés de référence afin d'identifier des unités textuelles appelées *SCU*s (Summary Content Units). Un

SCU est un ensemble d'unités textuelles (souvent des phrases) des résumés de référence encodant la même information. Un poids égal au nombre de résumés de référence qui l'instancient lui est assigné. Ces *SCU*s sont organisés en pyramide où les SCUs de même poids se trouvent à la même couche. L'évaluation d'un résumé automatique est effectuée en repérant les *SCU*s candidats qu'il contient. Par la suite, chaque *SCU* candidat hérite du poids du SCU le plus similaire dans la pyramide. Ensuite, le score *PYRAMID* est calculé comme le rapport de la somme des poids de tous ses *SCU*s candidats sur la somme des poids d'un résumé idéal ayant le même nombre de SCUs. Notez que le calcul du score *PYRAMID* a été automatisé en utilisant la sémantique distributionnelle (Passonneau, 2013). Cependant, l'annotation des résumés modèles reste une tâche ardue et difficile à automatiser.

2.3.3 BE

Au départ, chaque phrase des résumés de référence et du résumé à évaluer est réduite en ses éléments sémantiques minimaux (***BE***s : *Basic elements*). Ensuite, on fusionne tous les *BE*s de référence qui sont sémantiquement identiques et on attribue un score à chacun des BEs de référence. Puis, on compare chaque *BE* du résumé à la liste des *BE*s de référence et on calcule le score global de tous les *BE*s contenus dans le résumé candidat (Hovy, 2006). Notez que l'identification des *BE*s, se fait en utilisant des règles de découpage exécutées sur l'arbre syntaxique du texte d'entrée. La comparaison des *BE*s fait recours à des stratégies de comparaison telle que la similitude lexicale exacte, la similitude de lemmes, la similitude de synonymes, etc. Un point (1) est attribué à chaque BE du résumé candidat qui figure dans un résumé de référence (si et seulement si la similarité entre ce *BE* et un *BE* de référence est satisfaite). Le score global attribué au résumé généré est égal à la somme pondérée des poids de ses *BE*s.

2.3.4 Les techniques d’évaluation de résumés automatiques par abstraction

Les méthodes d’évaluation citées ci-dessus présentent deux problèmes majeurs : *i)* Ils ne permettent pas de mettre en œuvre des évaluations à large échelle, *ii)* les résumés de référence ne sont pas forcément idéaux si on tient compte de la subjectivité de la personne qui rédige un résumé (Hovy, 2006). De ce fait, plusieurs recherches ont essayé de proposer des méthodes d’évaluation entièrement automatiques (sans référence) qui permettent non seulement de réduire le coût de l’évaluation mais aussi de contourner le problème de la qualité des résumés manuels et de leur disponibilité. Dans ce contexte, plusieurs protocoles d’évaluation principalement basés sur des mesures de divergence ont été proposés. On en cite le système SIMetrix qui calcule la divergence entre la distribution de probabilité du vocabulaire du résumé produit et celle du texte source (Louis 2013). Fresa est un autre système qui utilise aussi des mesures de divergence (la divergence de Kullback-Leibler et celle de Jensen-Shanon) pour estimer la qualité d’un résumé généré automatiquement (Torres-Moreno, 2010). Les divergences peuvent être calculées, comme dans ROUGE, en employant différents types *de ngrammes*. Le système d’évaluation Fresa a été étendu dans la perspective du multilinguisme (Saggion et al., 2010). Toujours dans le même contexte, SummTriver (Cabrera-Diego 2016 and Cabrera-Diego, 2018) est un modèle d’évaluation qui calcule simultanément la dissimilarité entre trois éléments : le résumé candidat, le texte source et un ensemble de résumés automatiques de la même source mais produits avec des méthodes différentes. La divergence de Kullback-Leibler ou de Jensen-Shanon est utilisée pour calculer les dissimilarités entre ces éléments. Ensuite, elles sont fusionnées en utilisant la trivergence de distributions de probabilité (Torres-Moreno, 2015).

2.4 Conclusion

Dans ce chapitre, nous avons fait l'état de l'art des méthodes de résumés automatiques de textes et de méthodes d'évaluation de résumés générés automatiquement. Dans les pages qui suivent, nous présenterons une méthode de résumé automatique par extraction dans le troisième chapitre et une méthode cognitive de résumé automatique par abstraction dans le quatrième. Nous proposerons également une méthode cognitive d'évaluation des résumés de textes générés automatiquement au cinquième chapitre. Dans l'annexe, nous mettrons de l'avant un cas d'utilisation des résumés automatiques de textes comme moyen d'analyse et de synthèse de documents retournés par un système de recherche de documents basé sur l'outil LUCEN.

CHAPITRE III : RÉSUMÉ DU TEXTE ARABE VIA L'ÉQUILIBRAGE DU SAC À DOS DE LA RÉTENTION EFFICACE

3.1 Détails de l'article

ARABIC TEXT SUMMARIZATION VIA KNAPSACK BALANCING OF EFFECTIVE RETENTION

Alaidine Ben Ayed, Ismaïl Biskri and Jean-Guy Meunier

Article publié **Procedia Computer Science** d'Elseiver (Volume 189, pp 312-319, 2021).

Notre thèse s'inscrit dans le domaine du traitement automatique du langage naturel, et plus particulièrement dans le domaine du résumé automatique. L'objectif de cet article est de mettre de l'avant un modèle performant de résumé automatique par extraction. L'approche proposée permet d'identifier les phrases porteuses d'information la plus saillante. Elle est indépendante de la langue. Dans ce chapitre, nous testons ladite approche sur l'arabe, une langue sémitique qui présente beaucoup de défis en lien avec la morphologie et la complexité de syntaxe. L'approche proposée dans cet article servira de base pour une approche cognitive de résumé automatique de textes par abstraction, qui fera l'objet du prochain chapitre et qui sera testée sur des articles de journaux en anglais. Notez que le présent chapitre met l'accent sur la dimension purement informatique de notre recherche. La dimension cognitive des approches relatées sera expliquée en détail dans les deux chapitres qui suivront.

3.2 Résumé

Cet article présente un nouveau protocole de résumé automatique de texte arabe par extraction. Notre approche est basée sur l'équilibrage du sac à dos (*Knapsack Balancing* en anglais) de la rétention efficace. La rétention efficace fait référence à la

maximisation de la rétention tout en priorisant les thèmes saillants. Tout d'abord, les segments de texte sont mappés aux thèmes importants et un score de rétention efficace est calculé pour chaque phrase. Ensuite, la tâche de résumé est formulée comme un problème d'optimisation combinatoire. Le résultat final est généré par la maximisation de la rétention efficace. En d'autres termes, les phrases sélectionnées pour faire partie du résumé sont les meilleures pour coder les idées saillantes tout en couvrant la plupart des thèmes centraux. Les résultats expérimentaux montrent que l'approche proposée surpasse trois protocoles de résumé arabe de pointe.

3.3 Abstract

This paper presents a new extractive Arabic text summarization approach based on the Knapsack balancing of effective retention. Effective retention refers to maximizing retention while prioritizing salient themes. First, text segments are mapped to salient topics, and an effective retention score is computed for every sentence. Next, the summarization task is formulated as a combinatory optimization problem. The final output is generated through the maximization of effective retention. Aka selected sentences to be part of the summary are the best ones to encode salient ideas while covering most central themes. Experimental results show that the proposed approach outperforms three state-of-the-art Arabic summarization protocols.

3.4 Introduction

3.4.1 Automatic text summarization

Automatic text summarization (ATS) refers to the process of producing a concise, reliable, and fluent summary from a longer script (Widyassari, 2019 and El-Kassas 2021). In essence, it is intended to assist humans in consuming relevant massive information automatically. ATS approaches fall into many subcategories: depending on the adopted strategy to produce the summary, automatic text summarization techniques might be categorized into extractive vs. abstractive protocols. Extractive techniques roughly extract a subset of the most salient sentences present in the original text to construct the generated output. In contrast, the abstractive ones apply paraphrasing techniques to produce novel sentences while preserving the source text's conveyed information. Also, there are single-document vs. multi-document summarization techniques depending on the number of documents to summarize. Summary techniques might be indicative or informative depending on the output. Depending on generality, they might be generic or query-driven. Based on the language criteria, they might be monolingual or multilingual. Note that novel summarization protocols are greatly needed to process the ever-growing amount of textual data available online, especially for the Arabic language, which has many unique characteristics.

3.4.2 Arabic Language

Arabic is an ancient Semitic language written from right to left (Guellil, 2019). It is deemed to be over 2000 years old. According to (Al-Saleh, 2016), it serves more than 300 million people globally as their native language, and it is one of the current six official and working languages of the United Nations. Its alphabet is made of 28 letters whose shape is determined by their positioning. Unlike many other languages, where words continuously change their meanings, Arabic words stayed firm and pure to their roots for faraway decades due to the religious heritage. Arabic operates with a triliteral root system. For instance, the word سيارة meaning "Car," comes from the three root letter letters (س ي ر) meaning: "traveling through Movement." Therefore, a word with a similar root-like سار, "he took a walk," would have a similar meaning. Therefore, guessing the meaning of a problematic Arabic word can be a straightforward task since it can be broken down into its three root letters. This root system is partially present in other languages, but Arabic consistently sustains this rule throughout its entire language, making its richness even more exciting to explore.

There are two types of sentences in the Arabic language: verbal and nominal. Verbal sentences have at least a verb and a subject, and they usually begin with a verb. The subject, as well as the object, might be attached to the verb. For instance, in شاهدتك ("I saw you") the conjugated verb شاهد ("to see") has a subject and an object suffix pronoun attached to it. Nominal sentences have two parts: a subject or topic (مبتدأ), and a predicate (خبر). They usually begin with a noun or a pronoun. Arabic is a highly concise and rich language; a three-word English sentence corresponds to a five-letter word in Arabic. For instance, "He gave him" is أعطاه in Arabic. Also, Arabic has over 60 different words for love. Each one has its own slightly different shade of meaning.

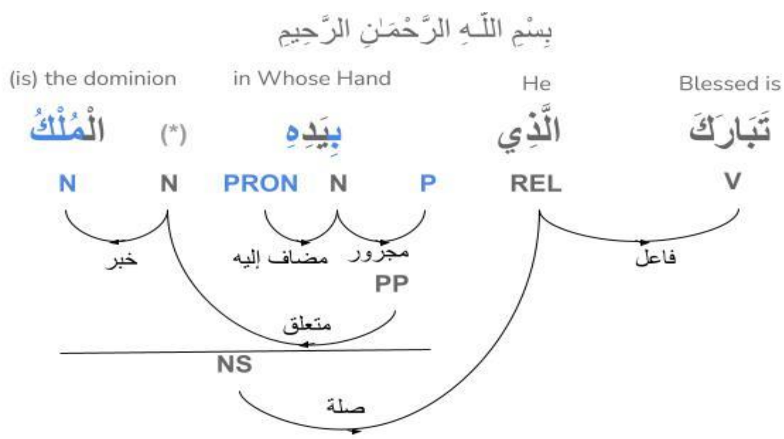

Figure 3-1: words organization in a sample Holy Quran Surate (Chapter 67). Surat Al-Mulk (Domination)

Dealing with the Arabic language implies many challenges (Al-Saleh, 2016), including:

- Arabic orthography: the shape of a letter depends on its position in a word. For example, the shape of the letter (ف), changes depending on whether it occurs in the beginning (فـ), middle (ـفـ), or end (ـف) of a word.
- The inflection in morphology: Arabic is highly derivational. Thus, there is a high inflection in morphology.
- Syntax’s complexity: It is due to how words are constructed and organized (Figure 3-1).

All the mentioned above issues make Arabic text summarization a challenging research area. In the next section, we make a literature review of the state-of-the-art Arabic text summarization techniques.

3.5 Literature review

(El-Kassas, 2021, Al-Saleh, 2016 and Al-Qassem, 2017) report the state of the art of most relevant Arabic automatic text summarization protocols. They fall into five main subcategories: discourse-based, graph-based, machine learning-based, ontology-based, and hybrid approaches.

Most discourse-based Arabic text summarization approaches lay on the Rhetorical structure and the segmented discourse representation theories. The Rhetorical Structure Theory (RST) is usually used to analyze rhetorical relations between text sentences. It considers a given sentence as a nucleus or a satellite. Nuclei segments contain essential information, while satellites provide additional information about the nucleus (Mann, 1988). A coherent text can be represented as an RST tree, and many variants of Arabic summarization protocols based on this intuition are proposed in (Al-Sanie, 2005, Mathkour, 2009, Mustafa, 2012, Ibrahim, 2013, and Maaloul, 2013). The Segmented Discourse Representation Theory (SDRT) is a dynamic semantic theory of discourse understanding and interpretation (Asher, 2005). It was employed in (Keskes, 2015) to propose an Arabic text summarization protocol.

Graph-based approaches lay the following hypothesis: the document to summarize can be described as an undirected graph. Every sentence is represented by a node. An edge between two nodes is activated if there is a relationship between their associated text units. A connection might be a similarity measure above a given threshold or any other type of relationship. Once the graph is set, the summary construction is based on analyzing sub-graphs of connected nodes encoding the most salient topics. The summarization approach proposed in (Al-Taani, 2014) is based on the mentioned above intuition. Furthermore, Arabic machine-learning-based approaches fall into two primary sub-categories; supervised vs. unsupervised methods. Supervised approaches apply machine learning techniques to learn patterns of a summary sentence and a non-summary sentence (Sobh, 2009, Abdel-Fattah, 2009 and Belkebir, 2015), while unsupervised methods use clustering techniques to generate the summary (Oufaida 2014, Fejer 2014, Haboush 2012 and EL-Haj 2011).

Ontology-based approaches, especially query-based Arabic summarization systems, use lexical databases to expand the user query to generate a relevant summary as a response. (Imam, 2013) uses the Arabic WordNet (AWN) (BLACK 2006 and Abouenour, 2013) for this purpose.

Finally, Hybrid Arabic text summarization approaches combine the mention above techniques. For instance, (Ahmed, 2013 and Azmi, 2012) add a meta-heuristics to the linguistic classic RST-based approach. Also, (Belguith, 2014) used an ontology to expand user queries, then used decision trees to construct query-based summaries. (Imam, 2013 and El-Haj, 2013) used Latent Semantic Analysis to map the query and the text to summarize onto a lower-dimensional space.

Notice that Discourse-based Arabic text summarization approaches have to overcome all the issues related to the Arabic language nature, such as identifying elementary discourse units' boundaries. Indeed, punctuation marks are not widely used in Arabic texts, so that that entire paragraph can be written without punctuation. Thus, relying on punctuation marks alone is not satisfactory for Arabic text segmentation. Also, the named entity recognition is not a straightforward task since Arabic does not have capital letters. Furthermore, Arabic discourse connectives are highly ambiguous. Also, Graph-based approaches fail to find out the best informative and salient text segments. Machine learning-based approaches ignore pertinent words only used in the test dataset. Those words are considered irrelevant since they were not seen during the training process. All mentioned above approaches, including hybrid ones, fail to maximize retention while prioritizing salient themes. According to (Lloret, 2012), determining the valuable features that excerpt salient ideas from the text to summarize while covering all central themes is the most challenging issue in extractive text summarization.

This chapter proposes a new Arabic text summarization approach based on the Knapsack balancing of effective retention. Effective retention refers to maximizing retention while prioritizing salient themes. The proposed approach considers the summarization task as a combinatory optimization problem. Text segments are mapped to salient concepts, and the summary is generated by maximizing of effective retention. The rest of this chapter is broken down as follows: the sixth section describes the proposed Arabic summarization approach. The seventh one compares the proposed approach to three state-of-the-art Arabic summarization techniques. The eighth section puts forth conclusions.

3.6 The proposed approach: KBER (Automatic Arabic text summarization via Knapsack balancing of effective retention)

3.6.1 Text preprocessing

Preprocessing stage involves four steps: *a)* text segmentation, *b)* tokenization, *c)* lexeme building, and d) stemming. First, the input text is broken down into unitary text units (sentences) during the segmentation phase. Next, sentences are split into tokens based on punctuation and white spaces. Afterward, the El-Khair Arabic stop-word list (El-Khair, 2017) is used to remove all stop-words. Initially, the lexeme is made of the remaining significant words. Next, the Information Science Research Institutes (ISRI) Arabic Stemmer (Larkey, 2002) is used to stemming text sentences and lexeme tokens. Only one instance of the same lexeme tokens is kept. By the end of the preprocessing stage, the text is coded as an $s \times t$ text feature matrix *TFM*. Note that s equals the number of segmented sentences, and t equals the number of lexeme tokens. Each line of *TFM* is a sentence feature vector ζ_i associated to sentence S_i. Each ζ_i is an IDF (Inverse Document Frequency) vector built using the generated lexeme.

3.6.2 Calculating the effective retention scores

Irrelevant and redundant information is mathematically expressed as a vector ω equal to the mean sums of sentence feature vectors ζ_i (Eq. 1). Furthermore, ω is used to generate a normalized sentence feature vector ζ_i for every text unit (Eq.2).

$$\omega = \frac{1}{s}\sum_{i=1}^{s} \zeta_i \tag{1}$$

$$\varrho_i = \zeta_i - \omega \tag{2}$$

Once we get rid of the redundant information by subtracting ω from each sentence feature vector, we build the basis of a more informative space. An SVD (Eq. 4) is performed on the covariance matrix $\aleph$ (Eq. 3) to construct a set of eigenvectors, the targeted space's unitary vectors. Those unitary vectors excerpt salient themes conveyed in the text to summarize.

$$\aleph = \frac{1}{s}\sum_{n=1}^{s} \varrho_n \varrho_n^T = \chi^{\mathrm{T}} \tag{3}$$

$$\chi = \delta\, \mathrm{S}\, \gamma^T \tag{4}$$

$\chi = [\varrho_1, \varrho_2, \dots, \varrho_s]$ (Eq. 3). $\aleph$ is a $t \times t$ matrix. χ is a $t \times s$ matrix. Additionally, dimensions of δ, S and γ are respectively $t \times t$, $t \times s$ and $s \times s$ (Eq. 4). Notice that δ and γ are orthogonal ($\delta\delta^T = \delta^T\delta = Id_t \; and \; \gamma\, \gamma^{\,T} = \gamma^{\,T}\gamma = Id_s$). Additionally:

1. γ's columns are eigenvectors of $\chi^T\chi$ and δ 's columns are eigenvectors of $\chi\chi^T$.
2. σ_k (Eigenvalues of $\chi\chi^T$ and $\chi^T\chi$) are squares of s_k (singular values of $\aleph$).

For $k > s$, eigenvalues σ_k of $\chi\chi^T$ are null. Their associated eigenvectors can be neglected ($s < t$), and matrix δ and S can be truncated. Respective dimensions of δ, S and γ (Eq. 4) become $t \times s$, $s \times s$ and $s \times s$. Now, the basis of the targeted more-informative space Π_K (Eq. 5) is built using the K eigenvectors δ_i, associated to the highest K eigenvalues.

$$\Pi_K = [\delta_1, \delta_2, \dots, \delta_K] \tag{5}$$

Next, normalized feature sentence vectors are projected onto the newly constructed space. They can be written as a linear combination of the K eigenvectors of Π_K (Eq. 6). The vector $\aleph_{\varrho_i}(k) = \delta_k^T \varrho_i$ in Eq. 7 provides coordinates of each sentence in that space.

$$\varrho_i^{proj} = \sum_k \aleph_{\varrho_i}(k)\, \delta_k \tag{6}$$

$$d_i(\delta_q) = |\delta_q - \varrho_i{}^{proj}| \tag{7}$$

Next, we compute the Euclidean distance between δ_i and any projected sentence onto the more-informative space (Eq. 7), and we construct the effective retention tensor (Figure 3-2).

1st eigenvector v_1	[4 0.07]	[8 0.21]]	[12 0.33]	[11 0.67]
2nd eigenvector v_2	[2 0.06]	[3 0.13]	[5 0.14]	[6 0.51]
3d eigenvector v_3	[6 0.16]	[4 0.29]	[7 0.34]	[8 0.63]
4th eigenvector v_4	[5 0.19]	[6 0.22]	[4 0.31]	[2 0.78]
5th eigenvector v_5	[7 0.15]	[5 0.22]	[6 0.46]	[3 0.61]

: Computed **Tensor** of the first five eigen concepts with a window of 4 sentences

[*i* , *d*] : *i* is a sentence index, *d* = distance(S_i, v_j) ; *j* = 1 … 5.

Figure 3-2: The effective retention tensor

In the illustrated example by Figure 3-2, the window size *w* used to construct the effective retention tensor is set to 4. The first line of the tensor provides the best four text units to encode the first-order theme. Their projected vectors onto the constructed salient space in Eq. 6 have the smallest distances to the first order eigenvector (associated with the highest eigenvalue in Eq. 7. Similarly, the tensor's fourth line gives the best four sentences to encode the fourth most central theme. Note that the order of a given sentence in a given window w depends on to which extent it encodes a given theme. For instance, the second sentence is the best sentence to encode the second most crucial theme, while the sixth one is the last best sentence to encode the same theme in a window of four sentences.

The built tensor is used to compute an effective retention score (*ER*) for each sentence: The intuition behind the *ER* score is the following: a given projected sentence onto the constructed salient space should encode as much as possible the most important themes conveyed in the text to summarize while taking into consideration the importance order of each theme. The *ER* coefficient for each sentence is defined by Eq. 8.

$$ER_{kw}(S) = \frac{1}{k}\sum_{i=1}^{k} \alpha_i \left[1 + \frac{1-\psi_i}{w}\right] \tag{8}$$

K is the number of eigenvectors encoding principal themes, $\alpha_i = 1$ if sentence S encodes the i^{th} theme. It is equal to zero in the opposite case. ψ_i is the rank of S in the i^{th} window of the *effective retention* tensor.

1.3.1 Generating the summary through Knapsack-balancing of effective retention scores

Given n-tuples of effective retention $< ER_1, ER_2, \dots, ER_n >$, and sentence length $< l_1, l_2, \dots, l_n >$, and $L > 0$, the goal is to determine the subset $T \subseteq 1,2,\dots,n$ of sentences that:

$$maximizes \sum_{i \in T} ER_i \tag{9}$$

$$subjectto \sum_{i \in T} l_i < L \tag{10}$$

The time complexity of the brute-force solution of this combinatorial problem is $O(2^n)$. The next section's reported results are obtained with our dynamic programming solution whose time complexity is around $O(n \times C)$. C is a constant.

3.7 Experiments and Results

3.7.1 Experimental settings

The EASC (EL-Haj, 2010) dataset and the ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) toolkit (Chin-Yew, 2004) are used for experiments. The EASC dataset contains 153 Arabic text article and their 765 human-generated extractive summaries. The average number of sentences per text to summarize is 17, and the average number of tokens is 377. Statistics about the number of articles per topic, average text length, and average sentence length are given in Table 3-1.

Table 3-1 : The EASC dataset

ID	**Topic**	**Articles**	**Average text length (Sentences)**	**Avg text length (Tokens)**
1	Art & Music	10	17	372
2	Education	7	14	356
3	Environment	34	17	417
4	Finance	17	14	264
5	Health	17	17	333

6	Politics	20	14	380
7	Religion	8	16	404
8	Science& Technology	16	13	393
9	Sport	10	17	331
10	Tourism	14	28	480

The ROUGE-1 and ROUGE-2 metrics are used in experiments to compare an automatically produced summary against a set of human-produced reference summaries. They refer to the overlap of unigrams and bigrams, the produced output, and their associated reference summaries. Obtained results are compared to those obtained by Arabic summarization protocols proposed in (Al-Abdallah 2017 and Jaradat 2016).

3.7.2 Results and discussion

Table 3-2: The proposed approach (KBER) VS. PSO, GA and HS ROUGE-1 and ROUGE-2 scores

Approach	ROUGE	Recall	Precision	F-measure
KBER	ROUGE-1	0.6113	0.6108	0.5614
KBER	ROUGE-2	0.4912	0.4961	0.4690
PSO	ROUGE-1	0.5444	0.5882	0.5532
PSO	ROUGE-2	0.4483	0.4814	0.4538
GA	ROUGE-1	0.5713	0.5658	0.5476

GA	ROUGE-2	0.4710	0.4597	0.4465
HS	ROUGE-1	0.5758	0.5686	0.5495
HS	ROUGE-2	0.4810	0.4679	0.4540

Table 3-2 describes the obtained ROUGE-1 and ROUGE-2 scores for the proposed approach and the three state-of-the-art Arabic summarization protocols. Obtained results prove that the proposed method outperformed the (Al-Abdallah 2017 and Jaradat 2016) summarization techniques. Indeed, mapping the original space onto a more salient space in a compressed way yields two results: First, the basis of the constructed space encodes salient themes. Second, it enables mapping each projected sentence into the constructed space to a set of dominant themes. Furthermore, using the tensor of distances between unitary vectors of the compressed space and normalized projected sentence feature vectors makes it possible to assign high scores of effective retention to text units encoding principal themes while considering the saliency degree of each one of them. Also, formulating the summarization task as a combinatorial problem gives the best balance between maximizing both retention and compression ratios while ensuring the best coverage of salient themes conveyed in the text to summarize.

3.8 Conclusion and future work

In this study, we proposed a new Arabic mono-document summarization protocol. The proposed approach computes an effective retention score for every sentence. Then, it generates the summary through retention maximization while ensuring covering all salient themes. The proposed approach can be applied to any Semitic language. It can also be applied to long scripts or collections of scripts. Notice that the main drawback of the proposed approach is that it only focuses on the retention criterium. As said in the previous section, the coherence is out of the scope of this work. Currently, we are implementing rhetorical analysis mechanisms to ensure text coherence. In such a case, the summarization process will be considered a multi-objective optimization problem tending to maximize retention and coherence while achieving a given compression ratio.

CHAPITRE IV : UNE APPROCHE COGNITIVE ABSTRACTIVE DE RÉSUMÉ AUTOMATIQUE DE TEXTE BASÉE SUR UN MODÈLE DE COMPRÉHENSION DE TEXTE ISSU DES TRAVAUX RÉALISÉS EN PSYCHOLOGIE COGNITIVE

4.1 Détails de l'article

A COGNITIVE ABSTRACTIVE APPROACH OF AUTOMATIC TEXT SUMMARIZATION BASED ON A COGNITIVE PSYCHOLOGY TEXT COMPREHENSION MODEL

Alaidine Ben Ayed, Ismaïl Biskri and Jean-Guy Meunier

Article soumis au **Cognitive Science** de Wiley (Décembre 2021).

L'objectif de cet article est de proposer un modèle de résumé automatique par abstraction performant. L'approche proposée repose sur une variante de la technique présentée dans le chapitre précédent pour identifier les phrases porteuses des informations les plus saillantes. Ensuite, une architecture d'apprentissage profond du type sequence2sequence, est utilisée pour générer les paraphrases qui constituent le résumé automatique par abstraction. Il est à noter que le présent chapitre met l'accent sur la dimension cognitive de notre recherche. Nous en expliquons les détails de l'implantation computationnelle du modèle de compréhension du texte, issu des travaux réalisés dans le cadre de la psychologie cognitive, sur lequel repose notre approche de résumé automatique de texte par abstraction. Nous détaillons en particulier la simulation du système de mémoire assez complexe ainsi que la mathématique et la simulation computationnelle des processus cognitifs qui interviennent pour la compréhension du texte.

4.2 Résumé

Ce chapitre met de l'avant un nouveau modèle cognitif de résumé de texte par abstraction. L'approche proposée est un système en deux étapes. Tout d'abord, les phrases du texte source sont mappées aux sujets saillants, et un score de saillance est calculé pour chacune. Ensuite, la tâche de résumé est formulée comme un problème de logique floue. Les phrases assurant une couverture et une fidélité maximales sont sélectionnées pour faire partie d'un pré-résumé. Ensuite, les phrases du pré-résumé sont reformulées à l'aide d'un transformeur T5. Les résultats expérimentaux montrent que l'approche proposée surpasse trois protocoles de synthèse de pointe.

4.3 Abstract

This chapter proposes a new cognitive abstractive text summarization model. The proposed approach is a double-stage system. First, text segments are mapped to salient topics, and a salience score is computed for every sentence. Next, the summarization task is formulated as a fuzzy logic problem. Sentences ensuring maximum coverage and fidelity are selected to be part of a pre-summary. Sentences of the first stage's output are rephrased using a T5 transformer. Experimental results show that the proposed approach outperforms three state-of-the-art summarization protocols.

4.4 Introduction

Text summarization refers to the process of producing a brief and accurate synopsis of voluminous text or collection of texts while focusing on text units that convey the most salient information and without deviating the overall meaning. Automatic text summarization (ATS) automatically removes redundant and insignificant information to construct shortened versions of lengthy documents, which could be a burdensome and costly process if done manually. Thus, there is a need to propose new text summarization approaches that automatically deliver insights of textual data overloading the digital space to deal with big data is high-volume, high-velocity, and high-variety information assets that demand innovative forms of information processing.

4.4.1 Automatic text summarization

There There are two broad approaches for automatic text summarization: extraction (Allahyari, 2017 and Andhale, 2016) and abstraction-based (Kumar, 2016). Extractive approaches gauge a weight for every sentence of the source text. Then, they concatenate the most salient sentences to generate the shortened version of the text to summarize. Extractive approaches fall into statistical and linguistic models. They generally compute a score that determines the importance of each sentence of the original text. Statistical models can be divided into three broad subcategories: frequency, feature-based, and machine learning-based approaches. Frequency-based approaches assume that essential words are more frequent than other words in a document. So, they use the basic term frequency (word probability) (Nenkova 2006, 2005) or inverse document frequency measures (Filatova 2004, Fung 2006 and Galley 2006) to assign scores to each sentence of the source document. Feature-based approaches determine sentence relevance by the presence of different features such as the presence of title/headline words, sentence position, sentence length, etc. (Gupta 2010). Machine learning approaches learn from training data patterns of "summary sentence" and "non-summary sentence" (Svore 2007, Burges 2005 and Hannah, 2014).

On the other hand, linguistic models assume that structure and coherence can be modeled through rhetorical relations. Those models employ discourse analysis techniques that establish a formal representation of the knowledge contained in the text (Barzilay, 1997 and Asghar, 2014).

Abstractive approaches are not limited to merely select and rearrange salient text sentences. They involve complex language modeling techniques to generate a fluent summary containing new sentences that cover core information and preserve the intended meaning of the source text. Thus, abstractive summarization is relatively more complex than extractive summarization.

There are many other taxonomies of automatic text summarization techniques based on different angles of view. For instance, a given summarization protocol may be mono-document or multi-document, depending on the number of input documents. It may be generic or query-based depending on whether the generated output is intended to report all text source events or focus on the user's searched topics. An automatically generated summary may also be evaluative if the summary is made subjectively or neutral in the opposite case.

4.4.2 Objectives and contribution

This chapter aims to bridge cognitive psychology with natural language processing to propose a dual-stage cognitive abstractive automatic text summarization protocol covering core information and preserving the author's intentions. The proposed approach relies on a cognitive psychology reading comprehension theory's computational model.

The following section reviews cognitive psychology models of text comprehension. The sixth one describes the proposed approach. Conducted experiments and obtained results are described in section seven. Conclusion and future work are exposed in the eighth section.

4.5 Related Cognitive psychology research on Reading comprehension

Reading comprehension has been a hot research topic since early 1978 (Kintsch, 1978). It has mainly focused on reading comprehension's cognitive processes ranging from recognizing letters and words to making predictions and inferences. For instance, the resonance Model focused on the reader's mental presentation (Myers, 1998). Since sand propositions may remain in the working memory, the resonance model pretends that the reader's mental presentation may be accessible in part while the reading process progresses. Secondary concepts are forgotten and maybe activated, reinstated by a sentence that is being read. The reinstatement process is either top-down or bottom-up. The top-down interpretation claims that readers tend to connect incoming text statements to earlier ones. When a connection between working memory and the text's mental representation fails to be established, the reader tends to reinstate his earlier working memory to find a link. The bottom-up interpretation is based on earlier research, which confirmed that a reader's mental presentation of a text depends on the elements of the sentence being processed (Albrecht, 1995). Aka words of the sentence being read activate previous statements when reinstating them to the working memory. The Landscape model mimics how those prominent text items are being initiated, saved, and retention rose in memory to construct a moderately steady memory representation of a text (Van Den Broek, 1996).

Both resonance and landscape models incorporate causality in their assumptions without simulating it. The causality-based effects simulation is the center of focus of the Langston and Trabasso model (Trabasso, 1999, Langston, 1999) since statements highly relying on previous story events are customarily read faster (Myers, 1987) and further often rated as relevant to the text (Trabasso, 1985).

The construction-integration model claims that the reading comprehension process should be decorticated beyond relationships between explicitly stated statements in the text (Kintsch, 1991). It focuses on the inferencing subprocess, which generates new knowledge based on the information being processed. Also, it may bring saliant background knowledge into the reader's subconscious thoughts. In this case, text propositions would retrieve a set of elements from the reader's world knowledge net during the construction phase. Then, appropriate ones are selected during the integration phase. Furthermore, the construction-integration model assumes that knowledge is processed at three different mental representation levels during the reading comprehension process: *i)* a literatim design enabling recognition of words, *ii)* a semantic one that encodes preeminent text items, and *iii)* a situational design of the situation to which the text endures. Note that the Gestalt model suggested an alternative view of the construction-integration model's external world knowledge claiming that the knowledge is created by gathering event sequences in a microworld (St. John 1992a, 1992b).

Subjectivity is out of the scope of the mentioned above models. It was addressed by the Predication model (Kintsch, 2001). Also, previously mentioned approaches are localist models, aka text items and relationships are modeled separately. However, the Predication model employs a distributed representation of words and propositions; there is no borderline between them. Therefore, text items are described by vectors encoding relations between them. Mathematically, the Predication model relies on LSA (Latent Semantic Analysis) model discourse units (Landauer, 1998). The Golden and Rumelhart model (Golden 1993, 1994) also dealt with subjectivity issues. As mentioned previously, in the construction-integration model, an inference results from a search process through the reader's world knowledge. One drawback of this assumption is that the reader's world knowledge is subjectively defined. To overcome this problem, Golden and Rumelhart view inference as a form of pattern completion (Golden 1993, 1994).

Dealing with the text in the situational level of the construction-integration model calls for concepts or propositions that originate from the reader's knowledge and not from the processed text (Kintsch, 1978, 1988, 1991). So, the construction-integration (C-I) model focuses on knowledge instead of text. To overcome this drawback of the C-I model, the Distributed Situation Space Model (DSS) shares Gestalts's assumption claiming that text events take place in a microworld to proven the focus on the text at hand and make the amount of knowledge to be implemented manageable (Frank, 2003). Also, the Distributed Situation Space, which gets its name from the distributed nature of the space, holds on most architectural and mathematical assumptions the Golden and Rumelhart Models (Golden 1993, 1994).

Other cognitive psychology models of comprehension, like the Structure Building model, studied the involved processes in comprehending various media other than texts like pictures (Gernsbacher, 1995). They can be extended to deal with more complicated tasks like video comprehension. The Structure Building model splits the comprehension process into three sub-actions: *i)* building the base (foundation) of the text's mental representations, *ii)* mapping knowledge onto the constructed base, and *iii)* shifting the new structures when a piece of new information is not incongruity with the existing ones, or when dealing with new ideas.

4.6 CogSum: a new cognitive abstractive summarizer

The proposed summarization technique is a double-stage system. First, an extraction phase is performed. Extracted sentences should satisfy maximum retention and fidelity criteria. Next, the by extraction selected sentences are paraphrased using a Text-To-

Text Transfer Transformer. Note that the proposed system relies on Kintsch's construction-integration model of text grasp, assuming that Knowledge construction and knowledge integration are the two phases of text comprehension. During the construction phase, the mental representation of the discourse is modeled through a complex propositional network made of nodes and connections, meant to reflect any relationship between the unitary discourse elements (text sentences). Next, an elaborated propositional network is constructed. It encodes most of the salient knowledge and hidden concepts. In other words, the comprehension process refers to activating salient knowledge during the construction phase. The integration process refers to the spread of this activation of salient concepts and marginal ones' deactivation across the network. Note that when we read a text, our mind processes it at three levels:

- The surface structure level
- The intermediate level (the textbase)
- The cognitive level: the situation model

The surface structure level is simply the text's words and how they relate to each other at a syntactic level. The second level of information processing is the textbase, in which a proposition codes a basic unit of text. Each proposition refers to a given idea (or concept). Finally, the situation model (SM) integrates basic meanings derived from the textbase into our knowledge.

Figure 4-1 describes the proposed cognitive summarization protocol and its different information processing levels inspired by the construction-integration model of text grasp. Next, we detail the construction and integration phases.

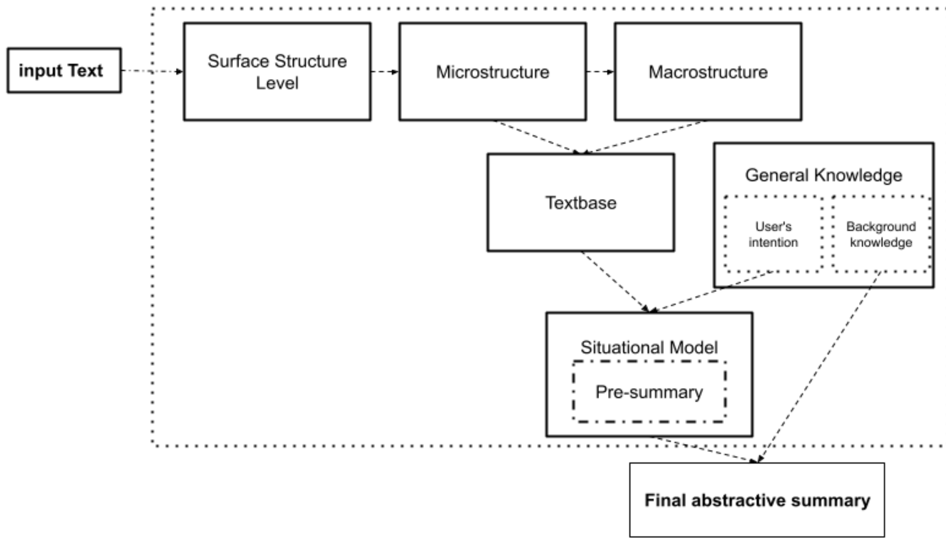

Figure 4-1: The proposed cognitive summarization protocol and its different information processing levels inspired by the construction-integration model of text comprehension

4.6.1 The construction phase

Mathematically, we project the source document onto a lower-dimensional space that captures the essence of concepts present in the source text. The latter space's unitary vectors are used to compute *retention-fidelity* scores as follows: First, a lexicon, including all unique non-generic words, is constructed. Each text unit (sentence) S_i is encoded by a sentence column feature vector x_i of d components. Each component of x_i corresponds to the number of occurrences of a given lexicon word (Eq. 1). Next, sentence feature vectors are stacked as rows of a data matrix to construct the crude text feature matrix; the microstructure (Eq. 2). Next, we will build the elaborated propositional network (the macrostructure). The latter will be used later to compute a saliency score for each sentence. Thus, the mean sentence vector is computed as described in Eq 3. It is subtracted from each sentence feature vector to remove noise and redundant information, and the normalized text feature matrix is constructed by stacking zero-centered sentence feature vectors as its rows (Eq 4).

$$x_i = \begin{pmatrix} w_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix} \text{ and } x_i^T = (x_{i1}, x_{i2}, \ldots, x_{id}) \tag{1}$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix} \tag{2}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}\ , \quad \ldots , \quad \frac{1}{n}\sum_{i=1}^{n} x_{id}\right)^T \tag{3}$$

$$X = \begin{pmatrix} x_{1-\mu^T}^T \\ x_{2-\mu^T}^T \\ \vdots \\ x_{n-\mu^T}^T \end{pmatrix} \tag{4}$$

As said previously, the goal is to project the initial sentence feature vectors dataset from many correlated coordinates (the microstructure) onto fewer most salient and uncorrelated ones called principal concepts (the macrostructure). Thus, the covariance around the mean is computed as described in Eq 5:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T = \frac{1}{n-1}\ X^T X \tag{5}$$

Vectors encoding those concepts will be built sequentially in a way that maximizes their contributions to the variances of the original set of sentence feature vectors. Mathematically, the goal is to find a collection of $k \leq d$ unit vectors $v_i \in \mathbb{R}^d$ (for $i \in 1, \ldots, k$) called principal concepts, such that: The variance of the set of sentence feature vectors projected onto the v_i direction is maximized. v_i should be orthogonal to $v_1, \dots, v_{i-1}$. The projection of a vector $x \in \mathbb{R}^d$ onto the line determined by any v_i is simply given as the dot product $v_i^T x$. The variance of the sentence feature vector x projected onto the first principal concept $v_1$is defined as follows:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (v_1^T x_i - v_1^T \mu)^2 = v_1^T S v_1 \qquad (7)$$

To construct v_1, S is maximized while satisfying the $\|v_1\| = 1$ additional constraint. The Lagrange multipliers (*LM*) approach is used to solve this optimization problem. *LM* implies that $Sv_1 = \lambda_1 v_1$,aka; v_1 is an eigen concept (mathematically, it is an eigenvector of the covariance matrix S). Note that $\| v_1 \| = v_1^T v_1 = 1$, this means that the corresponding eigenvalue is equal to $v_1^T S v_1 = \lambda_1$. It equals the variance of the sentence feature vectors along v_1. The most important concept is coded by the eigenvector associated to the highest eigenvalue. Next, the sentence feature vectors set is projected onto a new direction v_2 , the same way, while satisfying the $v_1 \perp v_2$ condition, then onto v_3 while satisfying $v_3 \perp v_1, v_2$, and so on.

By the end of this process, the first k vectors encoding principal concepts (macro-propositions) of X are built. They are eigenvectors of the covariance matrix S corresponding to its k highest eigenvalues. Next, the macrostructure will be constructed such that the k most important eigen concepts (macro-propositions) will form its orthonormal basis Ξ_k:

$$\Xi_k = [v_1, v_2, \dots, v_k] \qquad (8)$$

Each normalized projected sentence onto the constructed conceptual space can be written as a linear combination of k eigen concepts. Ξ_k is the macrostructure.

4.6.2 The integration phase

Next, the goal is to build a retention-fidelity tensor. Thus, the Euclidean distance between a given concept v_j ; $j = 1, \dots, k$ and any normalized sentence $\widehat{x_i} = x_i - \mu$, projected in the macrostructure is defined and computed as follows:

$$d_i(v_j) = \|v_j - \widehat{x_i}\| \tag{9}$$

1st Eigen Concept v_1	[3 0.09]	[6 0.19]]	[12 0.32]	[4 0.66]
2nd Eigen Concept v_2	[5 0.07]	[3 0.11]	[4 0.13]	[6 0.47]
3d Eigen Concept v_3	[6 0.18]	[4 0.33]	[7 0.37]	[9 0.75]
4th Eigen Concept v_4	[5 0.22]	[6 0.24]	[7 0.29]	[2 0.65]
5th Eigen Concept v_5	[12 0.17]	[5 0.47]	[6 0.48]	[3 0.59]

: Computed **Tensor** of the first five eigen concepts with a window of 4 sentences

[*i* *d*] : i is a sentence index, d = distance(S_i, v_j) ; j = 1 … 5.

Figure 4-2: Retention-Fidelity tensor construction using the five most important macro-propositions (eigen concepts)

The *Retention-Fidelity* tensor (Figure 4-2) provides distances between algebraic sentence feature vectors and the orthonormal conceptual space (macrostructure) basis's unitary vectors. It is constructed such that the line order depends on the importance of a given macro-proposition, while the column order is related to the extent to which a random sentence encodes a given macro-proposition. For instance, the first line provides the *w* best sentences to encode the first most crucial macro-proposition (their normalized projected feature vectors have the smallest distances to v_1 encoding the most important macro-proposition). The second line provides the same information related to the second most important macro-proposition, and so on. Note also that the fifth sentence, for instance, is the best sentence to encode the second most crucial macro-proposition, while the sixth sentence is the last one in a window size of four sentences. Next the Retention-Fidelity tensor will be used to compute a Retention-Fidelity score for each sentence.

Retention-Fidelity (RF) score computation and summary construction

First, a *Retention* score is computed for each normalized sentence being projected onto the constructed macrostructure. A given sentence having a high *Retention* sore should encode as much as possible the most important concepts (macro-proposition) expressed in the source document. In other words, it should appear as much as possible in a window of size w while taking into consideration the k macro-propositions. Mathematically, it is defined as follows:

$$R_{kw}(s) = \frac{1}{k}\sum_{i=1}^{k} \alpha_i \tag{11}$$

$\alpha_i = 1$ if the sentence S occurs in the i^{th} window. If not, it is equal to zero.

Now, an extended fidelity *(*$F_{kw}(s)$*)* score is computed for every sentence. It is a kind of averaged sum of the *retention* coefficient. The latter one is weighted according to the sentence's position in each window of size w. The central intuition is that sentences with a high F_{kw} score should encode important concepts (macro-propositions) while focusing on the most important ones. The *fidelity* score is defined as follows:

$$F_{\ kw}(s) = \frac{1}{k}\sum_{i=1}^{k} \alpha_i \left[1 + \frac{1 - \psi_i}{w}\right] \tag{12}$$

$\alpha_i = 1$ if a sentence s occurs in the i^{th} window. If not, it is equal to zero. ψ_i is the rank of a sentence s in the i^{th} window. Next, we use fuzzy logic to compute a unified retention and fidelity scores (*R-F*) as we proceeded in (Ben Ayed, 2019). The integration phase involves activating text units that encode the macrostructure's salient concepts (macro-propositions) to create a situational model (a pre-summary). The reader's intention (maximizing retention and fidelity) guides this activation process; Sentences with the highest R-Fs are chosen to be part of the pre-summary.

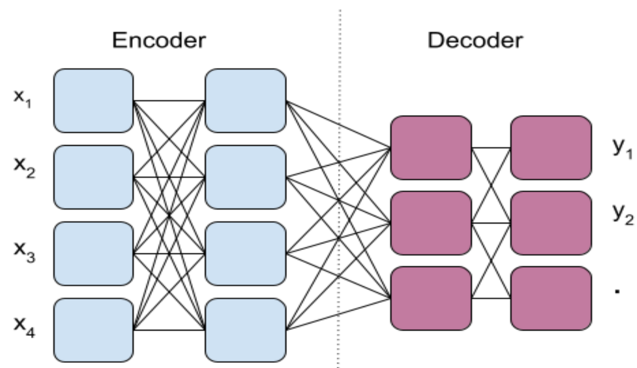

Figure 4-3: The used text-to-text-transfer-transformer architecture for sentence paraphrasing

Next, the **T5** model (**text-to-text-transfer-transformer**), proposed by **(**Raffel, 2019**)**, is used to paraphrase sentences of the generated pre-summary. It has an encoder (trained in a BERT manner) and a decoder (trained in a GPT manner). BERT is a Masked Language Modeling (MLM) model. Training Bert to predict masked tokens is performed as follows: *1)* Corrupting; Adding noise by replacing a random subset of the input with a mask token, *2)* Denoising; predicting the original tokens for each of the masked tokens. The Bert attention calculation for a given token depends on all remaining tokens in the sequence. The GPT model has proven well suited to text generation tasks. Its attention calculation for a given token only depends on the tokens that occur before it in each sequence. The T5 (text-to-text-transfer-transformer) model is essentially getting the best of both Bert and GPT worlds. The T5 model is pre-trained on the C4 dataset for the corrupting, denoising (Raffel, 2019). This pre-training encapsulates/simulates the reader’s background knowledge. The final output is a cognitive abstractive summary.

4.7 Experimental Results:

4.7.1 Dataset

The Timeline17 dataset is used for experiments (Giang, 2013). It consists of 17 manual-created timelines and their associated news articles. They mainly belong to 9 broad topics: BP Oil Spill, Michael Jackson Death (Dr. Murray Trial), Haiti Earthquake, H1N1 (Influenza), Financial Crisis, Syrian Crisis, Libyan War, Iraq War, Egyptian Protest. Original articles belong to news agencies, such as BBC, Guardian, CNN, Foxnews, NBCNews, etc. Those news articles are in plain text file format and noise filtered.

4.7.2 Results and discussion

To evaluate our proposed system (CogSum) for automatic text summarization, we used the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric to compare the by-extraction generated pre-summaries to reference abstracts produced by three state-of-the-art automatic summarizers (Luhn, 1958, Mihalcea 2004b and Erkan 2004). ROUGE-n (n = 1,2) refers to the overlap of n-gram between generated and reference summaries. ROUGE-S refers to Skip-bigram based co-occurrence statistics, while Skip-bigram is any possible pair of words in their sentence (Here the order does not matter). Obtained results are reported in In Table 4-1.

Table 4-1: Evaluation of the proposed system for automatic text summarization by comparing it to the Luhn, TextRank and LexRank approaches.

Summarizer	Rouge-1	Rouge-2	Rouge-S
Luhn	0.119	0.018	0.022
TextRank	0.204	0.206	0.047
LexRank	0.042	0.047	0.037
CogSum pre-summaries	0.241	0.061	0.058

Obtained results show that the proposed system outperforms summarizers proposed in (Luhn, 1958, Mihalcea 2004b and Erkan 2004). Indeed, our model can project the text onto a lower-dimensional space. The basis of this new space encodes the most salient information expressed in the source document. Extracting sentences with the highest R-F scores guarantees that both retention and fidelity criteria are met.

Figures 4-4, 4-5, and 5-5 illustrate a paraphrased output sample. The first one reports a BBC news article dealing with Michael Jackson ‘s death. The second one reports the generated pre-summary, and the final one reports the final abstractive cognitive generated summary.

Michael Flanagan of the DEA describes the operation Police have searched the Las Vegas home and offices of Michael Jackson 's doctor as part of a manslaughter investigation into the singer 's death. Dr Conrad Murray 's lawyer, Edward Chernoff, said officials were looking for the star 's medical records. The search is the second in a week following a similar operation at the doctor 's Houston clinic on 22 July. Dr Murray, who was with Jackson and tried to revive him before he died, has not been named as a suspect. In a statement, Dr Murray 's lawyer Edward Chernoff said the warrant `` authorised investigators to look for medical records relating to Michael Jackson and all of his reported aliases ''. He added Dr Murray was present during the search of his home and assisted the officers, who seized mobile phones and a computer hard drive. Reports suggest the investigation around Jackson 's death is focusing on his use of powerful painkilling drugs. The Drug Enforcement Administration has been involved in the investigation because the agency licenses doctors to administer controlled pharmaceuticals. Searches at the clinic and another site rented by Dr Murray in Houston, Texas, were carried out last Wednesday after a warrant was issued by a judge in the city. Dr Murray has already been interviewed twice by police the warrant, filed in Harris County District Court, said authorities were looking for `` items constituting evidence of the offence of manslaughter that tend to show that Dr Conrad Murray committed the said criminal offence ''. Such charges against a doctor for the death of a patient are extremely rare and require authorities to show there was a reckless action that created a risk of death. Items seized during the searches included 27 tablets of the weight loss drug Phentermine, a tablet of the muscle relaxant Clonazepam, two hard drives, notices from the Internal Revenue Service and a registration for controlled substances. Police have said Dr Murray is co-operating in the investigation. Paramedics were called to Jackson 's Los Angeles mansion while Dr Murray was performing CPR on the singer on the day he died, according to a recording of a 911 call. Speaking a few days after Jackson 's 25 June death, Mr Chernoff, denied his client administered painkilling drugs that could have contributed to the singer 's death. An official determination of what killed Jackson will not be made until the results of a toxicology report are disclosed.

Figure 4-4: A BBC news article about Michael Jackson’s death (2009-07-29).

Michael Flanagan of the DEA describes the operation Police have searched the Las Vegas home and offices of Michael Jackson 's doctor as part of a manslaughter investigation into the singer 's death. Dr Murray, who was with Jackson and tried to revive him before he died, has not been named as a suspect. Reports suggest the investigation around Jackson 's death is focusing on his use of powerful painkilling drugs. Dr Murray has already been interviewed twice by police the warrant, filed in Harris County District Court, said authorities were looking for `` items constituting evidence of the offence of manslaughter that tend to show that Dr Conrad Murray committed the said criminal offence ''. An official determination of what killed Jackson will not be made until the results of a toxicology report are disclosed.

Figure 4-5: The generated pre-summary from the BBC news article

Michael Flanagan of the dea describes the police's raid on the las vegas home and offices of michael jackson's doctor as part of an investigation into his death. Dr Murray, who was with Jackson and tried to revive him before he died, has not been named as a suspect. reports suggest that the investigation is focused on jackson's use of powerful painkilling drugs for pain treatment. reports suggest that the investigation is focused on jackson's use of powerful painkilling drugs for pain treatment. Dr Murray has already been interviewed twice by police the warrant, filed in Harris County District Court, said authorities were looking for 'items constituting evidence of the offence of manslaughter that tend to show that Dr Conrad Murray committed the said criminal offence '. A formal determination of what killed jackson will not be made until a toxicology report is released

Figure 4-6: The final generated cognitive abstractive summary

The generated final summary can sometimes have minor issues. Some of the generated sentences are almost identical to the extracted ones, with only minor differences in a word or two. Additionally, incorrect, or awkward grammar can occur when dealing with complex sentences. Furthermore, the T5 model might not be as good on out-of-domain (from training data) inputs. The latter issue can be averted by using better training data.

4.8 Conclusion and perspectives

This chapter proposed a cognitive abstractive summarization approach. The proposed system bases on a cognitive psychology model of text comprehension. The first stage ‘s results show that the proposed cognitive summarization model outperforms three states of the art summarization techniques. Currently, we are testing a bunch of deep learning architectures to deal with the paraphrasing task. More specifically, we are implementing a BART autoencoder. The standard BART implementation performs training by *1)* applying a mask on a random subset of the input sequence (corrupting), *2)* learning a model to reconstruct the original text. We are testing a double stage paraphrasing technique which uses a sentence permutation corruption schema in the first stage (The input is split based on periods (.), and the sentences are shuffled), and text infilling corruption schema in the second one (Some text spans are each replaced with a single mask token).

CHAPITRE V : UN MODÈLE D'INTELLIGENCE ARTIFICIELLE EFFICACE ET EXPLIQUABLE D'ÉVALUATION DE RÉSUMÉ GÉNÉRÉS AUTOMATIQUEMENT : UN CAS D'UTILISATION POUR COMBLER PSYCHOLOGIE COGNITIVE ET LINGUISTIQUE COMPUTATIONNELLE.

5.1 Détails de l'article

AN EFFICIENT EXPLAINABLE ARTIFICIAL INTELLIGENCE MODEL OF AUTOMATICALLY GENERATED SUMMARIES EVALUATION: A USE CASE OF BRIDGING COGNITIVE PSYCHOLOGY AND COMPUTATIONAL LINGUISTICS

Alaidine Ben Ayed, Ismaïl Biskri and Jean-Guy Meunier

Article publié comme chapitre du livre **« Explainable AI Within the Digital Transformation and Cyber Physical Systems** », édition Springer (Octobre 2021, 2021ISBN 978-3-030-76408-1).

Nous avons proposé une méthode de résumé automatique par extraction et une approche de résumé automatique par abstraction dans les articles des troisième et quatrième chapitres respectivement. L'objectif de cet article est de présenter une méthode d'évaluation des résumés générés automatiquement par extraction. L'approche en question est la première ébauche de protocole cognitif d'évaluation de résumés automatiques de textes. L'objectif principal est d'avoir recours au modèle computationnel proposé dans le chapitre IV du modèle « *construction-integration* » de compréhension de textes, issu des travaux réalisés en psychologie cognitive, pour évaluer automatiquement la qualité d'un résumé généré par extraction.

5.2 Résumé

Le résumé automatique de texte (RAT) (un sous-ensemble du traitement du langage naturel) est activement intégré dans de nombreuses applications industrielles, telles que les systèmes de questions/réponses, l'analyse des textes juridiques ou le résumé de génération de titres de l'actualité. L'explosion de la quantité de données textuelles provenant de diverses sources dans le contexte de l'ère du big data et de la révolution industrielle 4.0 exige que des techniques innovantes de résumé de texte automatique, autrefois considérées comme inaccessibles, soient mises en œuvre dans un monde de plus en plus numérique. Ce chapitre se concentre sur une sous-tâche de RAT, à savoir l'évaluation des résumés générés automatiquement (ERGA). Il propose une approche cognitive et explicable de l'ERGA. Le modèle proposé s'appuie sur la version computationnelle du modèle Kintsch pour la compréhension de textes. Il a été testé et comparé à la méthode ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) : une approche standard pour évaluer automatiquement les résumés générés. Les expériences menées montrent que les scores de notre approche sont fortement positivement corrélés à ceux de l'état de l'art ROUGE. La contribution saillante de cette recherche est qu'elle propose un modèle cognitif et explicable d'ERGA, et qu'elle pousse vers le traitement cognitif du langage naturel. Elle démontre également comment la psychologie cognitive peut servir de base pour proposer une approche d'intelligence artificielle explicable (XAI) en justifiant les scores d'un modèle d'ERGA.

5.3 Abstract

Automatic text summarization (ATS) (a subset of natural language processing) is actively being integrated across many industrial applications, such as question answering, legal texts or news summarization, and headline generation systems. The

explosion in the amount of text data from various sources in the context of the big data and 4.0 industrial revolution era demands innovative automatic text summarization techniques, formerly thought unattainable, to be accomplished in an ever more digital world. This chapter focuses on a sub-task of ATS, namely automatically generated summaries evaluation (AGSE). It proposes a cognitive and explainable approach of AGSE. The proposed model is the computational version of the Kintsch model of reading comprehension. It was tested and compared to Recall-Oriented Understudy for Gisting Evaluation (ROUGE): a standard approach to automatically evaluate generated summaries. Conducted experiments show that our approach's scores are highly positively correlated to the state-of-the-art ROUGE ones. This research's salient outcome is that it proposes a cognitive and explainable model of AGSE, and it pushes toward cognitive natural language processing. It also demonstrates how cognitive psychology can be used for an explainable artificial intelligence (XAI) approach to justify an AGSE model's scores.

5.4 Introduction

5.4.1 Automatic Text Summarization

Automatic summarization has been adopted for the daily running of affairs. Book abstracts on digital bookstores, show trailers, and headlines on TV broadcasts are samples of summaries we deal with regularly (Widyassari, 2019 and El-Kassas, 2021). automatic summarization has commonly been defined as the process of condensing a piece of media form to a shorter version while preserving key informational elements (Radev, 2002). The spectrum of its application ranges from texts to audio and video media forms. The particular case of automatic text summarization (ATS) refers to creating a concise, reliable, and fluent abstract from a more extended reference text (Gambhir, 2017).

Following technological improvements, an enormous volume of textual records is publically accessible (Liang, 2017, Lee, 2017, Roetzel, 2019 and Schmitt, 2017). This massive amount of the available data calls for automatic text summarization, enabling access to only relevant information. (Torres-Moreno, 2014**)** argues that automated summarization has concerns deserving addressing despite having been a target of academic research for more than five decades. Also, it states six main arguments why we need automated text summarization. First, abstracts lessen the time spent on reading a more extended text. They make it possible to consume content efficiently. Second, they facilitate the selection process when searching for a document. Third, automatic summarization can likewise make the indexing process more effective when dealing with massive textual databases. Fourth, it provides us with less biased summaries than those made by humans. Fifth, automatically generated abstracts carry much personalized information, which can be a valuable supplement to question answering frameworks. Finally, automatic text summarization increases the number of texts that can be processed by commercial abstract services.

Depending on the angle of perception, there are many taxonomies of automatic text summarization (Mani, 2001). One critical criterion to consider when analyzing ATS approaches is the type of the generated output. The latter can be either an extract or an abstract of a source document. Extractive summarization implies that the original text's most significant segments are excerpted to make the abstract. On the flip side, abstractive summarization uses paraphrasing techniques to present the original format's significant issues logically. It produces the original summaries; that is why it is a more challenging task than extractive summarization. The number of documents to summarize is another criterion that categorizes the summarization process into mono-document and multi-document ATS. When taking language as an angle of view, we can distinguish three variants of ATS: *1)* mono-lingual automatic text summarization, when the source input and the final output are in the same language; *2)* multi-lingual automatic summarization, when the original text is written in more than one language, thus, the final output would be in the corresponding languages; and *3)* cross-lingual automatic text summarization, when the generated summary is not in the same language of the source text. The authors in (Saggion, 2013) have pointed out key challenges associated with automatically generated summaries evaluation.

5.4.2 Evaluation protocols of automatically generated text summaries

Significant advances in the AGSE research area have been made during the last two decades. Various evaluation protocols have been proposed in this context. Furthermore, many evaluation campaigns have been led since early 1996. SUMMAC (the TIPSTER Text Summarization Evaluation) (Mani, 2002), DUC (Document Understanding Conference) (Over, 2007), and TAC (Text Analysis Conference) (TAC, 2008) are the most far-reaching ones. Notice that the evaluation process can be carried out in reference to a human-made summary. It can also be conducted without an ideal reference (Lin, 2003).

Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is the standard metric for automatically generated summaries evaluation purposes (Lin, 2004). It compares the generated output to a set of reference human-produced summaries. There are five main variants of the ROUGE metric:

- ROUGE-N (Lin, 2004): captures the n-gram overlap between the input and output texts; for instance, ROUGE-1 refers to the overlap of unigrams between system and summary references. ROUGE-2 refers to the overlap of bigrams.
- ROUGE-L (Lin, 2004): measures the longest matching sequence of words using LCS. LCS does not require consecutive matches since it uses in-sequence matches that reflect sentence-level word order. In this case, there is no need to

fix a predefined n-gram length since LCS automatically includes the longest in-sequence common n-grams.

- ROUGE-W (Lin, 2004): is a bunch of weighted LCS-based statistics that hold serial LCSes.
- ROUGE-S (Lin, 2004): is a skip-gram co-occurrence metric. It captures any pair of words in a sentence in order, allowing for arbitrary gaps. For instance, skip-bigram measures word pairs' overlap with a maximum of two gaps between sentence tokens.
- ROUGE-SU (Lin, 2004): is a set of skip-bigram plus unigram-based co-occurrence statistics.

Ramirez-Noriega et al. (Noriega, 2018**)** proposed a new variant of the ROUGE protocol that does not involve human-built model summaries (ASHuR). ASHuR checks whether the most informative sentences of the original text were extracted. Informative sentences are selected based on the frequency of concepts they encode, the presence of cue words, and sentence length. RETENTION is another metric of automatically generated summaries evaluation (Hovy, 2005). It has been used in DUC evaluations. It gives insights on to which extent the extracted summary conveys critical information present in the source text. RESPONSIVENESS has also been used in focus-based summarization tasks of DUC and TAC evaluation campaigns (Conroy, 2008). It uses a 5-point ranking scale, indicating how well the summary satisfied many predefined information criteria. The PYRAMID evaluation is another approach that was built upon the same intuition. It uses SCUs (Summarization Content Units) to compute a set of weighted scores (Nenkova, 2004). An automatically generated summary containing units with higher weights would have a high PYRAMID score. An SCU weight for a given text unit is relative to its frequency in the human-made summaries. FRESA is another approach that does not involve human-produced reference summaries (Torres-Morenoa, 2010). It computes a set of divergences among probability distributions.

Another evaluation of text summaries without human references approach was recently proposed by Jonathan et al. (Rojas, 2021). It is based on the linear optimization of content metrics using a genetic algorithm.

Lloret et al. (Rojas-Simón, 2018) give an overview of challenging issues related to summary evaluation research that remains an effortless task. Notice that all of the previously proposed AGSE models were designed according to a classical natural language processing view that involves computer science, mathematical, and linguistic backgrounds. In this chapter, we present a new cognitive and explainable protocol of AGSE. The proposed approach relies on a reading comprehension model that emerged from cognitive psychology research. In the next section, we review the most critical cognitive psychology models of reading comprehension.

5.4.3 Cognitive psychology models for text comprehension

Several theories about reading comprehension have been proposed since early 1973 (Van Dijk, 1983). The proposed models tend to analyze different cognitive processes involved in the reading comprehension activity, including recognizing letters and words and getting their meaning, syntactic parsing of sentences, making predictions and inferences, etc. Below, we make a short review of the most crucial cognitive psychology models of text comprehension.

The Resonance model

The reader's mental presentation is the center of focus of the resonance model. It may be accessible in part as the reading progresses. This intuition is exemplified by the fact that sand propositions may remain in the working memory since they are essential to the text, and those concepts considered secondary are forgotten. The latter ones may be reinstated through reactivation: being instigated by a sentence that is being read.

Reinstatement is either top-down or bottom-up. The top-down interpretation precedes the argument that readers try to establish a relationship between incoming text statements and earlier ones. When a connection cannot be established between the working memory and mental representation of a text, it calls for earlier reinstatement of the reader's working memory to ensure a link. The bottom-up interpretation claims that there is nothing like active search processes. Hence, "elements from current sentences activate previous statements when reinstating them to the working memory." The latter assumption has been affirmed by an earlier study, which found out that a reader's mental presentation of a text is prone to resonate with the elements of a sentence being processed (Albrecht, 1995). The latter finding led to the establishment of the resonance model (Myers, 1998).

The Landscape model

The landscape model focuses on the construction of a relatively stable memory representation of a text, which is an essential facet of the comprehension process (Van Den Broek, 1996). The landscape model simulates how prominent text items are being activated, stored, and retention strengthened in memory.

The Langston and Trabasso model

It has been argued that statements having a robust causal relation to previous story events are usually read faster (Myers, 1987). They are further often recalled as well as rated as relevant to the text (Trabasso, 1985). It has further been argued that when a story is read, the reader can relate it to causal events. The resonance and landscape models have been found to incorporate causality in their explanations. However, they do not simulate it. The Langston and Trabasso model simulates all of the causality-based effects (Trabasso, 1999, Langston, 1999).

The Construction-Integration model

The construction-integration model is one of the most influential reading comprehension theories (Kintsch, 1991). It simulates many cognitive processes ranging from recognizing words to constructing a representation of text elements. Kintsch assumes that readers build three different mental representations: (*i*) a literatim representation of the manuscript, (*ii*) a semantic one that illustrates the essence of the text, and (*iii*) a situational representation of the situation to which the manuscript holds. The construction-integration model treats the reading comprehension process as much more than just relationships between explicitly mentioned information printed in the text. It casts light on the inferencing sub-process, which either brings relevant background knowledge into someone's subconscious thoughts or generates new knowledge based on what was read.

The Predication model

The predication model was designed to address issues relating to subjectivity (Kintsch, 2001). It employs a distributed representation of words as well as propositions. The main idea is to represent discourse as a network made up of connected nodes. The nodes represent discourse items, and the connections give insights into relations between them. Generally, in a localist model, items and relations are presented separately, while in a distributed representation, there is no clear line between them. Therefore, items are represented as vectors, which determine the relations between them. The predication model relies on Latent Semantic Analysis (Landauer, 1998) to automatically acquire an objective vector representation of discourse units.

The Gestalt model

Gestalt's model offers an alternative view of Kintsch's external world knowledge that the world of knowledge is formed by accumulating experiences of event sequences in a microworld. The model has proposed that world knowledge is an amalgamation of experiences in a microworld. The model has been criticized for harboring two issues. One, there is no representation of the order of story events. Two, the processing of a story propositional way requires an equal number of computations, which paves the way for the lack of processing time (St. John 1992a, 1992b).

The Golden and Rumelhart model

Inferencing is one of the most critical sub-processes involved in reading comprehension. Usually, it involves the reader's general knowledge to activate (retrieve) not explicitly mentioned information in a text. The abovementioned models differ in their view toward inferencing. In the construction-integration model, text propositions typically retrieve a set of associated propositions from the reader's world knowledge net. Then at the integration phase, the propositions that are considered most appropriate to the text are selected. Here inference is taken as a result of a search process through the reader's world knowledge. However, it has been identified that one setback for the construction-integration model is the subjectivity involved in defining the world knowledge net that is included in the model. To overcome this problem as well as issues related to the order of story events and the Gestalt model's processing time, Golden and Rumelhart view the inference as a form of pattern completion. Even though it seals the loopholes created by Gestalt's model, it has been criticized for involving a switchback from the distributed representation to the localist one (Golden 1993, 1994).

The Distributed Situation Space model

Kintsch and Welsch (Kintsch, 1991), Kintsch and Dijk (Kintsch 1978), and Kintsch (Kintsch, 1988) argued that there are three levels of text: a surface text level, the textbase level, and the situational level. Considering texts at situational levels has been criticized as it focuses on knowledge instead of text. Comprehension of texts calls for concepts or propositions that originate from the reader's knowledge and not from the text that is being processed.

Gestalts and Distributed Situation Space Models have similarities in that the amount of knowledge to be implemented is made manageable by letting stories take place in a microworld and that situations are represented distributively. Also, the Distributed Situation Space and the Golden and Rumelhart Models share most architectural assumptions and their mathematical basis from which it follows how world knowledge concerning relations between storytime steps is implemented and how the knowledge is applied to the story representation to result in inferences. It is also important to note that both models take the issue of situation space very seriously, and it is from the distributed nature of the space that the Distributed Situation Space model gets its name (Frank, 41).

The Structure Building model

The structure building model focuses on casting light on the involved processes in the comprehension of various media such as texts and pictures (Gernsbacher, 1995). It divides the comprehension process into three broad sub-processes: *a)* setting a foundation (base) for the text's mental representations, *b)* mapping information onto that base, and *c)* shifting the new structures when dealing with new ideas or when new information is not incongruity with the existing one.

5.4.4 Originality of Our Work

The salient outcome of this research is that:

- It proposes a cognitive protocol of AGSE since it is built upon a cognitive psychology model of reading comprehension.
- The proposed AGSE protocol gives insights on to which extent criteria of a good summary are met instead of merely focusing on the N-gram overlaps between the original text and the generated output.

Most of the summary evaluation protocols described in Sect. 5.4.2 were designed according to a pure classical natural language processing view that involves computer science, mathematical, and linguistic backgrounds. This chapter presents a new cognitive and explainable AGSE model. The proposed approach relies on a reading comprehension theory that emerged from cognitive psychology research. Furthermore, classic AGSE protocols only focus on N-gram overlaps between the original text and the generated summary. They do not give any insight on to what extent the criteria of a fair resume are met, namely:

- Retention (coverage): The generated output should cover all the concepts reported in the source document.
- Fidelity: The summary should accurately reflect the author's point of view by focusing on salient concepts conveyed in the original text.
- Coherence: The generated summary should be semantically meaningful.

Since previously proposed AGSE protocols merely focus on the N-gram overlaps between the original text and the generated summary, they only reflect on the retention ratio. They cannot check whether the fidelity criterion is met or not; If a newspaper article reports events related to five concepts and a given automatically generated summary focuses on the three marginal ones. It gets a higher relevancy score than another overview focusing on the two most crucial concepts present in the source text despite it is merely focusing on nonessential events.

This chapter presents a cognitive evaluation protocol of automatically generated text summaries. The proposed approach casts light on to which extent both retention and fidelity are met. The technical and mathematical details of the proposed AGSE protocol are detailed in the next section. The conducted experiments and obtained results are reported in the sixth section. Conclusion and future work are exposed in the seventh section.

5.5 CATSE: A cognitive automatic text summarization evaluation protocol

5.5.1 The main idea

The construction-integration (C-I) model of text comprehension comprises two ordered steps: knowledge construction and knowledge integration. During the construction phase, the C-I model generates a propositional network made up of nodes and connections that encode a crude mental representation of the discourse. The connections are meant to reflect any relationship between the discourse elements (Figure 5-1). Next, we will consider text sentences as elementary discourse elements.

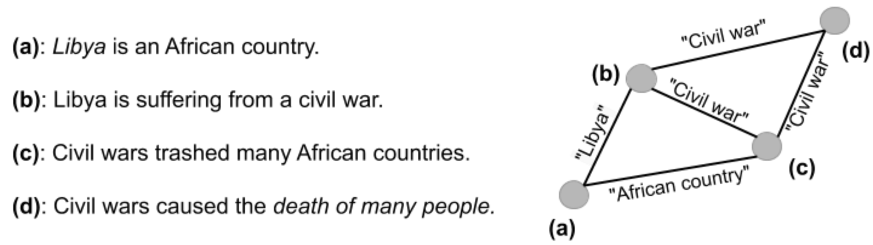

Figure 5-1: An example of expressing the crude mental representation of the discourse as a propositional network.

The mental representation is not refined at this stage. An elaborated propositional network is needed to interpret salient knowledge and infer hidden concepts. In the above example, the elaborated propositional network should reflect, in a condensed way, that the civil war is the dominant concept. It should also reveal the hidden semantic link between sentences *(a)* and *(d)*. In other words, it should help to infer that "many people are dead in Libya because of civil war." In this way, the mind is stimulated as a network, and the comprehension process refers to activating salient knowledge. This activation begins in the construction phase. The integration process refers to the spread of this activation of salient concepts and marginal ones' deactivation across the network (Figure 5-2).

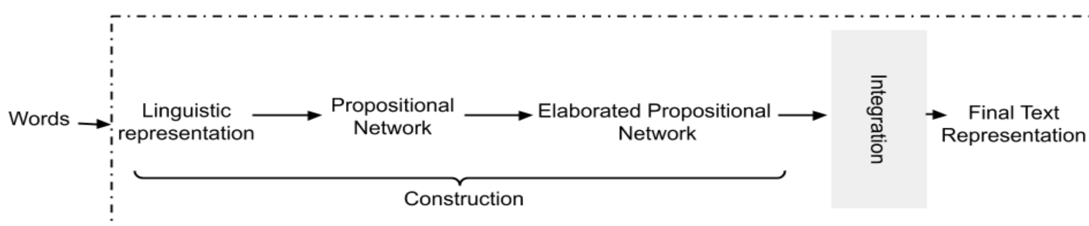

Figure 5-2: The whole text comprehension process according to the construction-integration model

The propositional network's set of propositions and their associated connections (semantic links) form what we call the microstructure (Figure 5-3). According to [39], the CRUD semantic representation of the text being read (the microstructure made up of micro-propositions) is too complex to be manipulated and memorized. Thus, the human brain tends to use a set of strategies (macro-rules) that aim to build a more abstracted semantic structure (the elaborated propositional network) containing the text's gist and known as the macrostructure (Figure 5-3). The macrostructure, made up of macro-propositions, is better suited for memorization or for manipulations that operate on it due to cognitive constraints like the memory size and the complexity of representations.

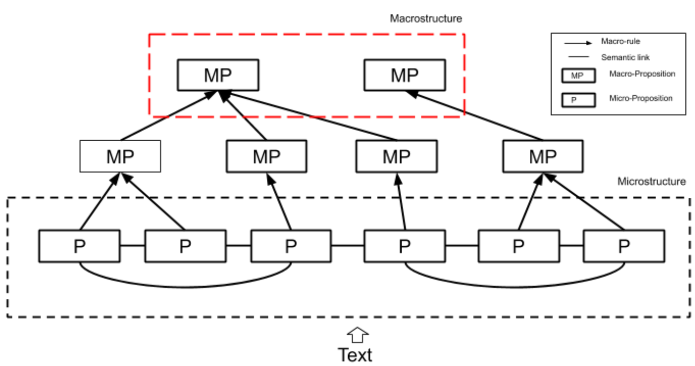

Figure 5-3: Application of the Macro-rules to generate the macrostructure from the microstructure.

From a computational perspective, the proposed CATSE protocol generates a feature vector for each input text sentence. Feature vectors encode the different concepts stated by their associated sentences. This representation is better than the propositions-based encoding proposed by the original construction-integration model as it provides semantic relations between text units. Also, it makes it possible to represent a text unit as a linear combination of concepts. The microstructure is made up of a set of feature vectors encoding text units. Latent Semantic Analysis is used to construct the macrostructure, the compressed form of the microstructure containing salient concepts stored in the semantic memory. Building the macrostructure is a mathematical transformation that consists of mapping feature vectors of the microstructure onto a lower-dimensional space whose unitary vectors encode critical concepts stated in the text. The macrostructure will be stored in the episodic memory.

The integration phase consists of activating text units that encode the salient concepts of the macrostructure. The reader's intention guides this activation process; the CATSE protocol mainly focuses on the retention and fidelity criteria. Coherence is out of the scope of this research. A score assessing the quality of the generated abstract will be produced by the end of the integration phase. It is equal to the average of the activation weights relating to the original text's sentences reported in the generated summary. Note that the construction-integration model of text comprehension (Kintsch, 88, 99) relies on three different memory structures:

- The semantic memory is simulated by LSA (Latent Semantic Analysis); the semantic similarity between text units is meant to model human associations in semantic memory.
- The working memory audits the mapping between salient concepts and their associated text units.
- The episodic memory keeps track of all concepts or propositions that occur in working memory and their activation values.

5.5.2 Levels of representation

When we read a text, our mind processes it at three levels:

- The surface structure level
- The intermediate level: the textbase
- The cognitive level: the situation model

Below we describe each level of representation, and we explain how we encode it in our summary evaluation protocol.

The surface level

The surface structure level is simply the text's words and how they relate to each other at a syntactic level. When we read a sentence, our mind initially understands it at a grammatical level. It checks whether words are in the correct order according to the grammatical rules of a given language. Also, it assesses the level of cohesion in the structure. At this superficial level of representation, our brain tries to understand the information conveyed by sequences of words. If it encounters a text in a language it does not know, it will reject it at the surface structure level. In other words, if we are not familiar with the sequence of terms printed in the text and their syntactic structure, our brain will not waste its cognitive processing time on them. The surface structure level can be considered the first "test" a printed text must go through to be deemed worthwhile for processing. Our CATSE protocol assumes that the summary to evaluate already passed this test since we deal with a by-extraction summary. Furthermore, the goal is to assess the quality of the generated summary, not the original text's quality.

The intermediate level: the Textbase

The textbase is the second level of information processing in which a proposition codes a basic unit of text. Each proposition refers to a given idea (or concept). Below we present an example of this from (Van Dijk, 1983).

(1) The fascists have won the election in El Salvador

(2) (i) FASCISTS (X_1)
(II) HAVE WON (X_1 , X_2)
(III) ELECTIONS (X_2)
(IV) IN (X_2 , X_3)
(V) EL SALVADOR (X_3)

Figure 5-4:Text unit codification in the construction-integration model (Van Dijk, 1983)

The example (Figure 5-4) illustrates a sentence *(1)* and its atomic propositions *(2)*. A complete proposition can replace any variable X_i. Therefore, (ii) could also be written HAVE WON (FASCISTS, ELECTIONS). The theoretical construction-integration model assumes that the whole text is transformed into atomic propositions before the next stage. The CRUD natural language is not suitable for computational processing since it serves many purposes other than the expression of meaning. Kintsch argues that: "propositions are designed to capture those semantic relations that are most salient in text comprehension" (Kintsch, 1988). One drawback of the construction-integration model is that it does not technically describe how those propositions would be generated from the CRUD text (Kintsch, 1988). Furthermore, when we deal with a long text, we generate many complex propositions, which would explode the processing time and affect the quality of the final output (Grabowski, 1992).

The central intuition is to break down the text into propositions to get into the essential meaning encoded by each unit (sentence) of the text. Thus, our CATSE protocol encodes each sentence by a feature vector whose coordinates are the *tf-idf* of relevant tokens (words) printed in the text. In this way, each text unit is encoded as a weighted linear combination of cue words (meaning concepts). As a result, instead of breaking down our sentence into n propositions, we encode it by a unique weighted feature vector. Semantic relations between text units can be assessed by computing cosine distance between feature vectors.

The cognitive level: the situation model

The situation model (SM) integrates basic meanings derived from the textbase into our knowledge. Kintsch claims that the SM is continually changing as we read. At any given time, it generally depends on someone's background knowledge and/or the reader's intentions when summarizing a text ranging from achieving given compression ratios to maximizing retention, fidelity, and coherence of the generated summary. In the coming sections, we will suppose that we are working in a closed reasoning word since the goal is to automatically assess the quality of the generated summaries, which are, by construction, supposed only to contain extracted sentences from the text to summarize. They are not supposed to integrate the reader's background knowledge since we are dealing with a "by extraction" summary evaluation protocol. We literally are in an encapsulated knowledge and rote memorization mode; thus, we only limit our focus on the automatic summarizer's intention to generate a fair summary. Modeling the comprehension processes that forge the situation model should be done at the semantic level, especially cognitive modeling. Thus, our CATSE protocol relies on LSA, a powerful model for representing the meaning of words and sentences.

5.5.3 The CATSE protocol of automatically generated summary evaluation

First, the text to summarize is segmented into units (sentences). Then, a lexicon is built and filtered to discard all universal expressions and terms. Next, the text is coded as an $s \times t$; s refers to the number of sentences, while t refers to the number of significant unique tokens. We initially map the propositional network (the microstructure) onto a random space during the construction phase. Then, we build the elaborated propositional network (the macrostructure). The latter will be used later to compute a score assessing the generated abstract's quality during the integration phase.

The construction phase

Each sentence S_i is coded by a sentence feature vector ζ_i of t components. ζ_i components refer to the *tf-idf*s associated to tokens present in a given sentence S_i. The set of ζ_i vectors encodes the micro-propositions and the feature text matrix, obtained by stacking sentence feature vectors ζ_i as its lines, encodes the microstructure. Afterward, redundant information is coded as ω; the mean sum of sentence feature vectors is ζ_i (equation 1). We normalize each sentence feature vector to excrete redundant information since the brain tends not to waste its cognitive processing time (Eq. 2).

$$\omega = \frac{1}{s}\sum_{i=1}^{s}\zeta_i \tag{1}$$

$$\varrho_i = \zeta_i - \omega \tag{2}$$

The macrostructure (the new space in which we map the elaborated propositional network) is built by first computing the covariance matrix described in Eq. 3. A singular value decomposition will then be performed as described by Eq. 4 to construct the macro-propositions (eigenvectors of $\aleph$ associated with the highest eigenvalues).

$$\aleph = \frac{1}{s}\sum_{n=1}^{n}\varrho_n\varrho_n^T = \varkappa\varkappa^T \tag{3}$$

$$\varkappa = \delta S\gamma^T \tag{4}$$

$\varkappa = [\varrho_1, \dots, \varrho_2]$ in Eq. 3. Also, $\aleph$ and $\varkappa$ are respectively $t \times t$ and $t \times s$ matrix. Also, dimensions of matrix δ, S and γ in Eq. 4 are, respectively, $t \times t$, $t \times s$ and $s \times s$. Note that, δ and γ are orthogonal ($\delta\delta^T = \delta^T\delta = Id_t$ and $\gamma\gamma^T = \gamma^T\gamma = Id_s$). Additionally:

- Eigenvectors of $\varkappa^T\varkappa$ are columns of γ.
- Eigenvectors $\varkappa\varkappa^T$ are columns of δ.
- Eigenvalues σ_k of $\varkappa\varkappa^T$ and $\varkappa^T\varkappa$ are squares of singular values s_k of S.

Eigenvalues σ_k of $\varkappa\varkappa^T$ are null when $k > s$ and their associated eigenvectors are unnecessary since $s < t$. So, matrix δ and S can be truncated, and, dimensions of δ , S and γ in Eq. 4 become, respectively, $t \times s$, $s \times s$, and $s \times S$. Next, the macrostructure Π_k will be built using K eigenvectors δ_i (macro-propositions), belonging to the highest K eigenvalues as shown in Eq. 5:

$$\Pi_k = [\delta_1, \delta_2, \dots, \delta_k] \tag{5}$$

The construction-integration theory claims that the construction stage refers to *(1)* building the microstructure (the CRUD semantic representation of the text being read), and *(2)* transforming it into a macrostructure coding the text's gist. The microstructure and the macrostructure form the textbase, the second level of knowledge processing by our minds. The shifting from the microstructure to the macrostructure is performed by applying a mathematical transformation that maps the original space in which we projected the initial propositional network onto a compressed, more relevant space (the macrostructure) whose unitary vectors are the macro-propositions (constructed vectors that better encode salient concepts yarned by sentences of the text to summarize). The mathematical transformation simulates the macro-rule that aims to build a more abstract semantic structure.

The integration phase

The integration phase consists of activating text units (sentences) that encode the macrostructure's salient concepts. The reader's intention guides this activation process; the CATSE protocol mainly focuses on the retention and fidelity criteria. Our CATSE protocol's primary concern is to assess the quality of the generated abstract. In other words, it will compute a score that approximates to which extent sentences of the original text are covering all the concepts conveyed in the source text while focusing on the most salient ones. Thus, sentence feature vectors are projected onto the constructed macrostructure and encoded as a linear combination of K macro-propositions as described by Eq. 6: the vector $\aleph_{\varrho_i}(k)\delta_k$ provides coordinates of a sentence S_i in the conceptual space.

$$\varrho_i^{proj} = \sum_k \aleph_{\varrho_i}(k)\delta_k \tag{6}$$

Next, the Euclidean distance between a given macro-proposition m and any projected sentence onto the macrostructure is defined and computed as described by Eq. 7:

$$d_i(\varrho_m) = \left\| \varrho_m - \varrho_i^{proj} \right\| \tag{7}$$

Next, we construct the Retention-Fidelity tensor (Figure 5-5) as follows: First, we fix a W window size. W is proportional to the compression ratio. In the below example, W is set to 4.

First macro-proposition	[1 0.09]	[16 0.12]	[3 0.45]	[8 0.48]
Second macro-proposition	[16 0.14]	[3 0.16]	[1 0.21]	[7 0.24]
Third macro-proposition	[4 0.21]	[7 0.32]	[8 0.43]	[3 0.5]
Fourth macro-proposition	[8 0.04]	[1 0.21]	[3 0.24]	[7 0.42]
Fifth macro-proposition	[7 0.5]	[6 0.52]	[4 0.69]	[16 0.92]

: Computed **Tensor** of the macro-propositions (salient concepts) with a window of 4 text units

[*i* *d*] : i is a text unit index, d = cosine distance between a text unit and a macro-proposition (salient concept).

Figure 5-5:The retention-fidelity tensor

The tensor's first line gives the four text units having the smallest distances to the vector, encoding the first macro-proposition (the most salient concept). The second line shows the same information relative to the second most important concept (macro-proposition). Note here that the order of a given text unit in a given window W depends on its cosine distance to a given macro-proposition. For instance, the first sentence is the best one to encode the first most salient concept, while the eight sentence is the last one to encode it in a window of four text units. Also, the first macro-proposition is encoded by the eigenvector related to the highest eigenvalue. Thus, it encodes the most salient concept. The second macro-proposition encodes the second most salient concept, and so on. The retention-fidelity tensor simulates the working memory. It will be used later to infer a unified fuzzy retention-fidelity score for each sentence of the source text: First, a retention score is computed to each text unit projected onto the macrostructure. It is defined as the number of times it occurs in the retention-fidelity tensor divided by the number of macro-propositions to be bounded between 0 and 1. The central intuition behind it is that a given sentence having a high retention sore should encode as much as possible of the K macro-propositions of the macrostructure.

$$R_{kw}(s) = \frac{1}{k}\sum_{i=1}^{k}\alpha_i \tag{8}$$

$\alpha_i = 1$ if s occurs in the i^{th} window. If not, it is equal to zero.

Next, we compute a fidelity score, defined as shown in the ninth equation, as the averaged sum of summary sentences' retention coefficients. The fidelity score's central intuition is that text units whose fidelity score is high should encode the most salient concepts stated in the source text. So, they should have minimum distances from the macro-propositions in Eq. 7. In other words, the fidelity score gives insights on to which extent a given sentence encodes concepts present in the original text (the macro-propositions) while taking into consideration the salience degree of each one of them. Mathematically, the fidelity score is defined as follows:

$$F_{\ kw}(s) = \frac{1}{k}\sum_{i=1}^{k}\alpha_i\left[1 + \frac{1-\psi_i}{w}\right] \tag{9}$$

$\alpha_i = 1$ if s occurs in the i^{th} window of the retention-fidelity tensor. If not, it is equal to zero. ψ_i is the rank of s in the i^{th} window.

Next, we use fuzzy logic to compute a unified retention-fidelity score (*R*-*F*) for each sentence of the source text. Text units having the highest (*R*-*F*) scores will be activated by the end of the integration phase. They will remain in the episodic memory, and they will present candidate sentences of an ideal summary. We opt for the fuzzy logic to compute the unified retention-fidelity score because the brain is a "fuzzy machine." Linguistic variables are input and output variables in simple words (Figure 5-6).

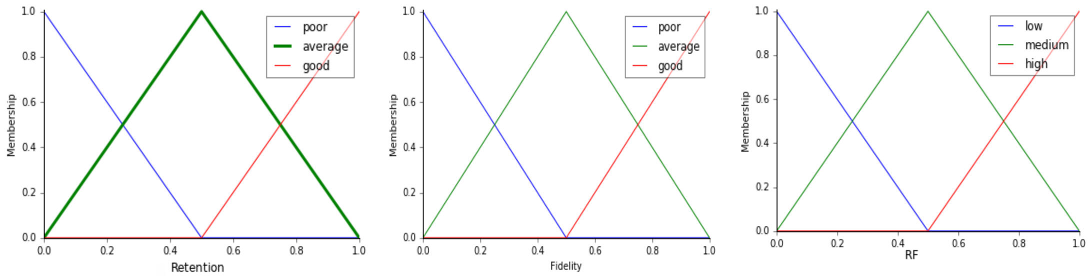

Figure 5-6: Defining linguistic variables and membership functions.

"Low," "medium," and "high" are the linguistic terms used to model the retention and fidelity scores. Afterward, we build a set of rules into the knowledge base in the form of IF-THEN-ELSE structures:

- ***Rule 1:*** *If the retention score is high or the fidelity score is high, then, the* ***R-F*** *score is high.*
- ***Rule 2:*** *If the fidelity score is medium, then, the* ***R-F*** *score is medium.*
- ***Rule 3:*** *If the retention score is low and the fidelity score is also low, then, the* ***R-F*** *score is low.*

Next, fuzzy set operations evaluate the previously defined rules to infer the fuzzy values of *R-F* sores. Here, the used operations for "OR" and "AND" are "Max" and "Min," respectively. Afterward, we combine all evaluation results to form final fuzzy *R-F* scores (Figure 5-7).

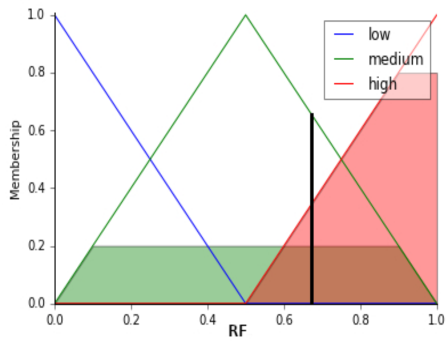

Figure 5-7: Fuzzy R-F score

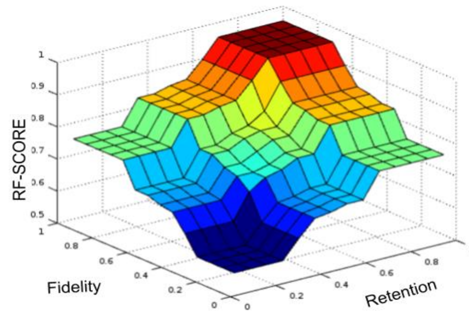

Figure 5-8: Defuzzification and activation of text units (sentences) having the highest *R-F* scores.

Defuzzification is performed according to the membership function for output variables, as shown in Figure 5-7. A unified retention-fidelity (*R-F*) score is computed for every sentence of the source text. Text units having the highest *R-F* scores will be activated (integrated into the situation model) and stored in the episodic memory. Sentences with low *R-F* scores will be deactivated and forgotten (Figure 5-8). The CATSE score is equal to the averaged sum of retention-fidelity (*RF*) scores of sentences chosen to be included in the actual summary.

5.6 Experiments and results

5.6.1 Datasets

In this chapter, we mainly used three datasets:

- The Timeline17 dataset (Giang, 2013): It consists of 17 manually created timelines and their associated news articles. They mainly belong to 9 broad topics: Original texts belong to news agencies, such as BBC, Guardian, CNN, Fox news, and NBC News. Major topics are: BP Oil Spill, Michael Jackson Death (M-J), Haiti Earthquake (H-E), H1N1 (Influenza), Financial Crisis (F-C), Syrian Crisis (S-C), Libyan War (L-W), Iraq War (I-W), and Egyptian Protest (E-P).

- Crisis dataset (EL-Haj, 2017): It consists of 20463 news articles dealing with the crisis in Egypt, Libya, Yemen, and Syria. Famous news agencies produced the original texts.
- EASC dataset (EL-Haj, 2017): Contains 153 Arabic articles and 765 human-generated extractive summaries of those articles.

5.6.2 Experimental results

To evaluate the proposed protocol of CATSE, we computed the Spearman correlation between ROUGE and CATSE scores using the three abovementioned datasets. Note that the Spearman correlation of two variables equals the Pearson correlation between their associated rank values: In contrast to the Pearson's correlation that assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not). A perfect Spearman correlation of +1 or -1 occurs when each variable is a perfect monotone function of the other. Those correlations were computed using reference texts and automatically generated ones using three summarizers:

- The Luhn summarizer (Luhn, 1958): Luhn's summarizer is a naive approach based on *tf-idf*. It extracts "salient" sentences of the source text. A given sentence's saliency depends on a bunch of meta-heuristics, including cue terms, sentence position in the text, and many other indicators.
- The TextRank summarizer (Mihalcea, 2004): TextRank builds an undirected graph using text units as vertices. The degree of semantic or lexical similarity between text units is attributed as a weight to vertices edges. The constructed graph is used to build a stochastic matrix. Next, the ranking over vertices is obtained by finding the eigenvector corresponding to the eigenvalue that gives a stationary distribution of the random walk on the graph.

- The LexRank summarizer (Erkan, 2004): It is also a graph-based summarizer. Measured semantic similarities or content overlaps between sentences are defined as weights to graph edges. LexRank uses the cosine similarity of *tf-idf* vectors in contrast to the TextRank approach that uses a very comparable weight based on the number of shared words between two sentences (generally normalized by the sentences' lengths).

Table 5-1: Spearman correlation between CATSE and ROUGE scores in the TIMELINE17 dataset

Sub-dataset	Support	Luhn	TextRank	Lexrank
BP-Oil	1415	0.74	0.66	057
E-P	563	0.81	0.62	0.64
H1N1	215	0.31	0.59	0.61
H-E	125	0.52	0.73	0.88
I-W	125	0.76	0.82	0.81
L-W	325	-0.16	0.23	0.19
S-C	214	0.82	0.86	0.91
M-J	124	0.79	0.83	0.86

Table 5-2: Spearman correlation between CATSE and ROUGE scores in the CRISIS dataset

Sub-dataset	Support	Luhn	TextRank	LexRank
Egypt	5110	0.72	0.66	0.64
Libya	5665	0.83	0.79	0.83
Syria	5355	0.51	0.81	0.82
Yemen	4333	0.74	0.69	0.39

The obtained results are reported in Tables 5-1, 5-3 and 5-3. Notice that if the Spearman correlation score is positive, it means that the two studied variables are positively correlated: if the first variable increases, the second increases. If the first variable decreases, it will be the same scenario for the second one. In contrast, if the Spearman correlation score is negative, it means that the two variables are negatively correlated. Also, if the Spearman correlation score is superior to 0.5, it means that the two variables are highly positively correlated. The obtained results from the 24 experiments led on the eight subsets of the Timeline17 dataset show that the ROUGE and CATSE scores are highly positively correlated in 20 scenarios. They are moderately positively correlated in three scenarios and negatively correlated in only in one scenario.

Table 5-3: Spearman correlation between CATSE and ROUGE scores in the EASC dataset

Sub-set	**Luhn**	**TextRANK**	LexRank
Art and music	0.11	0.42	0.16
Education	0.31	0.56	0.33
Environnement	0.33	0.51	0.52
Finance	0.29	0.61	0.74
Health	-0.21	0.02	0.013
Politics	0.16	0.29	0.24
Science and tehnology	0.52	0.49	0.32
Tourism	0.39	0.58	-0.25
Politics	0.53	0.50	-0.11
Sports	0.34	-0.21	0.41

The obtained Spearman correlation results using all subsets of the second dataset show that the ROUGE and CATSE scores are highly positively correlated. Better results are obtained when using long texts with many concepts. Our approach assumes that the source text to summarize reports at least two concepts. Thus, it is more suitable for long texts. Experiments conducted on the third dataset show that our method needs more parameter tuning to obtain a realistic assessment of the quality of generated summaries in Semitic languages.

5.7 Conclusion

This chapter focused on a sub-task of ATS, namely automatically generated summaries evaluation (AGSE). We proposed a cognitive and explainable approach of AGSE. The proposed model relies on the Kintsch theory of reading comprehension. Our evaluation protocol is tested and compared to Recall-Oriented Understudy for Gisting Evaluation (ROUGE): a standard approach used to evaluate automatically generated summaries. Conducted experiments on the Timeline17, Crisis, and EASC datasets show that our approach's scores are generally highly positively correlated to the state-of-the-art ROUGE ones. Best results are obtained when using long texts. A refined parameter tuning is needed when assessing the quality of generated summaries in Semitic languages. This research's salient outcome is that it proposes a cognitive and explainable model of AGSE, and it pushes toward cognitive natural language processing. It also demonstrates how cognitive psychology can be used for an explainable artificial intelligence (XAI) approach to justify an AGSE model's scores.

The current version of the proposed evaluation protocol only works with extractive summarization tasks. Now, we are implementing the abstractive oriented version of it. The main idea is the following: after identifying important sentences in the text, the next variant of our CATSE approach will tend to detect the reader's strategy to build the summary, which is viewed as applying adequate macro-rules. Macro-rules are the core of the cognitive processes involved in the summarization activity (Kintsch, 1978). Note that the construction-integration model states three types of macro-rules:

- Deletion: Each text unit containing minor redundant or unrelated details may not be considered part of the summary.
- Generalization: A generic segment may substitute a bunch of text units.

- Construction: Salient text units from retention and fidelity perspectives are stored in the episodic memory and activated later on to be part of an ideal summary.

The CATSE protocol should infer that a particular sentence in the text has been deleted given a text and its summary. In the same way, it should infer whether a sentence is a generalization, and so on. The selection (construction) macro-rule is already simulated using LSA: Each sentence of the summary is semantically compared with each source text sentence. A sentence of the source text will be considered deleted for the deletion macro-rule if no generated output sentence is sufficiently close to it. Similarly, a generalized sentence is a summary sentence close enough to more than one source text sentence.

CHAPITRE VI : CONCLUSION GÉNÉRALE

Tout au long de cette thèse, nous avons réalisé un tour d'horizon de techniques de génération et d'évaluation de résumés automatiques. Nous avons constaté que les approches proposées sont souvent purement mathématiques, linguistiques ou combinaisons des deux. Nous avons postulé au départ que le résumé par extraction ou abstraction ou l'évaluation d'un résumé automatique est une opération cognitive liée à des processus de compréhension de textes ou de discours et qu'on peut simuler algorithmiquement l'opération cognitive de compréhension de texte afin d'élaborer des modèles de résumés par extraction et abstraction ainsi que des protocoles d'évaluation de résumés automatiques de textes pertinents. L'étude des différents aspects lors de la production (humaine) d'un résumé et l'analyse des modèles proposés dans le cadre des travaux effectués en psychologie cognitive qui tentaient de décortiquer le processus de compréhension des textes nous ont permis de proposer de nouvelles approches cognitives de résumé automatique et d'évaluation de résumé automatique de textes.

6.1 Résultats

Dans le troisième chapitre, nous avons mis de l'avant un nouveau protocole de résumé automatique mono-document. L'approche présentée calcule un score de rétention effective pour chaque phrase. Nous avons défini la rétention effective comme consistant à « couvrir le maximum de thèmes tout en se focalisant les plus saillants ». La tâche de résumé automatique est traitée comme un problème d'optimisation du type « 0-1 *Knapsack balancing* ». Ainsi, le résumé est généré par la maximisation de la rétention tout en assurant la couverture de tous les thèmes saillants. L'approche proposée a été testée sur la langue arabe. Elle est applicable à n'importe quelle langue sémitique. Elle peut également être appliquée sur de longs textes ou à des collections de textes. Les résultats expérimentaux montrent que l'approche proposée surpasse trois protocoles de résumé arabe de pointe.

Dans le quatrième chapitre, nous avons mis de l'avant un modèle cognitif de résumé automatique par abstraction performant. L'approche présentée repose sur une variante de la technique proposée dans le chapitre III pour identifier les phrases porteuses des informations les plus saillantes. Ensuite, une architecture d'apprentissage profond du type sequence2sequence est utilisée pour générer les paraphrases qui constituent le résumé automatique par abstraction. Dans ce chapitre, nous avons mis l'accent sur la dimension cognitive de notre recherche. Nous avons expliqué les détails de l'implantation computationnelle du modèle de compréhension du texte, issu des travaux réalisés dans le cadre de la psychologie cognitive, sur lequel repose notre approche de résumé automatique de texte par abstraction. Nous avons détaillé en particulier la simulation du système de mémoire assez complexe ainsi que la mathématique et la simulation computationnelle des processus cognitifs qui interviennent pour la compréhension du texte. Notre méthode a été testée sur des articles de journaux comparativement à des systèmes de résumé automatique de référence. Les résultats reportés montrent que la nôtre atteint les meilleures performances.

Dans le cinquième chapitre, nous avons présenté une approche cognitive d'évaluation des résumés générés automatiquement par extraction. L'approche proposée est la première ébauche de protocole cognitif d'évaluation de résumés automatiques de textes. L'objectif principal était d'avoir recours au modèle computationnel proposé dans le chapitre IV du modèle « *construction-integration* » de compréhension de textes, issu des travaux réalisés en psychologie cognitive, pour évaluer automatiquement la qualité d'un résumé généré par extraction. Le modèle proposé a été testé et comparé à la méthode ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) : une approche standard pour évaluer automatiquement les résumés générés. Les expériences menées montrent que les scores de notre approche sont fortement positivement corrélés à ceux de ROUGE.

Dans l'annexe, nous avons exposé un cas d'utilisation des résumés automatique de texte comme outil d'analyse de documents retournés par un moteur de recherche. L'apport principal de ce travail, outre le fait de proposer un système original de recherche de document et une évaluation comparative de techniques originales d'expansion de requêtes, est qu'il a permis de tester notre approche de résumé automatique de texte dans un contexte multi-documents.

En guise de conclusion, la contribution saillante de cette recherche est qu'elle présente des modèles cognitifs de résumé automatique de texte ou d'évaluation de résumé automatique et qu'elle pousse vers le traitement cognitif du langage naturel. Elle démontre également comment la psychologie cognitive peut servir d'outil pour proposer des approches innovantes de TALN.

6.2 Perspectives

Plusieurs perspectives de recherche découlent de l'analyse d'approches de résumé automatique et d'évaluation de résumé automatique de texte proposées dans le cadre de cette thèse. Premièrement, l'article présenté dans le chapitre III se concentre uniquement sur le critère de rétention. Des mécanismes d'analyse rhétorique pourraient être ajoutées, comme on a fait dans (Ben Ayed, 2019) pour assurer la cohérence du texte. Dans un tel cas, le processus de résumé sera considéré comme un problème d'optimisation multi-objectifs tendant à maximiser à la fois la rétention et la cohérence tout en atteignant un taux de compression donné. Dans ce cas, la génération de résumé automatique peut être formulée comme un problème d'optimisation du type « *Knapsack Balancing* » bidimensionnel avec une contrainte d'équilibrage. En outre, le test de plusieurs architectures encodeurs-décodeurs ainsi que l'emploi de plus grands jeux de données mèneraient à une meilleure validation d'approche de résumé automatique par abstraction proposée dans le chapitre IV.

Une autre direction de recherche à approfondir serait de continuer le travail sur la version actuelle de notre approche cognitive d'évaluation de résumés automatiques pour qu'elle soit utile en l'absence de résumés de références. Pour ce faire, une fois qu'on a identifié les segments saillants (les plus informatifs), il faut déterminer la stratégie qu'un humain aurait élaborée pour construire le résumé produit, qui consiste à appliquer les macro-règles adéquates. Il est à noter que le modèle *construction-intégration* définit trois types de macro-règles :

- La suppression : chaque segment du texte contenant de l'information redondante ne devrait pas faire partie du résumé.
- La généralisation : un segment générique peut remplacer plusieurs autres segments.
- La construction : les segments les plus pertinents du point de vue de la couverture et de la fidélité seront enregistrés dans la mémoire épisodique pour qu'ils fassent partie d'un résumé idéal.

Notre prochaine version d'évaluation des résumés automatiques générés par abstraction devrait inférer les macro-règles mises en œuvre pour générer le résumé.

Une autre direction de recherche serait de proposer d'autres approches de résumés automatiques de textes et d'évaluation en ayant recours à d'autres modèles de compréhension et en travaillant sur des supports de données plus complexes que le texte. À titre d'exemple, le modèle de « *Structure Building model* » peut servir d'outil pour proposer une approche cognitive de résumé automatique de séquences vidéo.

ANNEXE : UN FRAMEWORK DE BOUT EN BOUT DE RECHERCHE DE DOCUMENTS/INFORMATIONS EFFICACE À BASE DE LUCÈNE

7.1 Détails de l'article

AN END-TO-END EFFICIENT LUCENE-BASED FRAMEWORK OF DOCUMENT/INFORMATION RETRIEVAL

Alaidine Ben Ayed, Ismaïl Biskri and Jean-Guy Meunier

Article accepté pour publication dans ***International Journal of Information Retrieval Research*** (IJIRR : Volume 12, Issue 2, Article 5.).

Nous avons proposé une méthode de résumé automatique par extraction, une approche cognitive de résumé automatique par abstraction et un protocole d'évaluation des résumés générés par extraction dans les articles des chapitres III, IV et V. Le présent article présente un cas d'utilisation des résumés automatiques de texte comme outil d'analyse de documents retournés par un moteur de recherche. Il est à noter que nous avons prouvé que notre approche de résumé automatique est invariante par rapport à la langue. Nous l'avons testée sur l'arabe (une langue sémitique) au chapitre III et l'anglais (une langue germanique) au chapitre IV. Toutes les expériences précédentes ont été réalisées dans un contexte de résumé automatique mono-document. Cet article, outre le fait de proposer un système original de recherche de document et une évaluation comparative de techniques originales d'expansion de requêtes, permet de tester notre approche de résumé automatique de texte dans un contexte multi-documents.

7.2 Résumé

Dans le contexte du big data et de l'ère de la révolution industrielle 4.0, il s'avère indispensable d'améliorer l'efficacité des systèmes de recherche de documents/d'informations pour gérer le volume toujours croissant de données textuelles dans un monde de plus en plus numérique. Cet article décrit un système en deux étapes de recherche de documents/d'informations. Tout d'abord, un outil de recherche de document, basé sur Lucene, est implémenté. Ensuite, quelques techniques d'expansion de requête utilisant un corpus comparable (*Wikipédia*) et le *word embedding* (« plongement de mots » ou « plongement lexical » en français) sont proposées et testées. Subséquemment, on procède au résumé automatique des documents récupérés pour créer un extrait court, précis et fluide des top documents retournés par notre système de recherche de documents. Les résultats obtenus montrent que le plongement lexical constitue un excellent moyen d'améliorer la précision d'un système de recherche de documents. De plus, les résumés obtenus satisfont les critères d'un extrait pertinent, à savoir la conservation et la fidélité.

7.3 Abstract

In the context of big data and the 4.0 industrial revolution era, enhancing document/information retrieval frameworks efficiency to handle the ever-growing volume of text data in an ever more digital world is a must. This chapter describes a double-stage system of document/information retrieval. First, a Lucene-based document retrieval tool is implemented, and a couple of query expansion techniques using a comparable corpus (*Wikipedia*) and word embeddings are proposed and tested. Second, a retention-fidelity summarization protocol is performed on top of the retrieved documents to create a short, accurate, and fluent extract of a longer retrieved single document (or a set of top retrieved documents). Obtained results show that using

word embeddings is an excellent way to achieve higher precision rates and retrieve more accurate documents. Also, obtained summaries satisfy the retention and fidelity criteria of relevant summaries.

7.4 Introduction

Document Retrieval (*DR*) is defined as the process of matching some stated user queries against a set of free-text records (Anwar, 2010 and Vishal, 2021). Nowadays, massive and quite variant data is being generated at an unprecedented rate. In this context, the big data era has overturned classical *DR* challenges. More focus is being addressed on proposing innovative indexing and searching routines. Document retrieval systems generally perform two basic operations: *1)* indexing; representing data in a condensed format, *2)* querying; is the process of querying the *DR* system to retrieve appropriate data. The first operation does not involve end-users. Generally, it is performed in an off-line mode. The second one includes numerous processing operations, ranging from filtering, searching, mapping to ranking returned indexes.

Document retrieval frameworks are built upon the cluster hypothesis (Fiana, 2013). Identifying the appropriate cluster of pertinent documents to a given straightforward user query is an easy task. Finding the clusters appropriate to complex queries is a more difficult task (Tombros, 2002 and Liu, 2006). The retrieval performance drops down if top accurate documents are not presented at the top of returned indexes. Proposing new ranking query-specific cluster strategies has been a hot research topic for many years (Leuski, 2001), and suggested solutions based on a cluster-against-query representation comparison (Liu, 2004 and Liu, 2008). Some document retrieval frameworks make use of extra features, including inter-cluster and cluster-document similarities (Kurland, 2006, 2008, 2011). Query expansion (*QE*) is another way to heighten document retrieval systems accuracy (Hiteshwar, 2019).

First attempts of query expansion have been proposed since early 1960. The main objective is to improve the retrieval process performance. In this context, *QE* was used as a procedure for literature indexing and searching (Maron, 1960). The user's feedback was employed in (Rocchio, 1971) to expand queries. (Jones, 1971 and Van, 1977) suggested a collection-based term co-occurrence *QE* protocol, while (Jardine, 1971) (Minker, 1972) introduced a cluster-based one. The mentioned above techniques led to satisfactory results. Nevertheless, they were experimented with using small corpora and a set of straightforward user queries. Researchers noticed a considerable loss in retrieval precision when the mentioned above techniques were tested using bigger corpora sizes provided by public search engines, firstly implemented in 1990 (Salton, 1990 and Harman, 1992). Consequently, query expansion has been a hot search topic, notably in an ever-growing big data word (Raza 2019). *Precision* and *Recall* are the states of the art standard measures of document retrieval accuracy (Sagayam, 2012). The first one refers to the percentage of relevant retrieved records, while the second one refers to the percentage of relevant records being retrieved. Notice also that the document retrieval research community uses *TRECEVAL*, a standard tool commonly used to evaluate ad hoc retrieval runs, given the returned documents and a conventional collection of refereed results.

Automatic text summarization (*ATS*) is another critical research area related to text document retrieval if we assume that the returned result may be a concise, reliable, and fluid extract of a given longer retrieved text document. *ATS* can also be applied to a set of retrieved documents. Generally, automatic text summarization is either performed by extraction or abstraction (Widyassari 2020). The first approach extracts prominent sentences that vehicle the essential concepts of the source text. Nevertheless, the latter creates novel sentences by applying rephrasing techniques instead of merely reporting the most salient fragments.

Note that extractive models gained more attention than abstractive approaches. Extractive summarizers generally estimate a relevancy score for each sentence of the original retrieved source text document. The latter score determines to what extent a given sentence encodes significant concepts. The generated output is made of the top-scored sentences. This kind of summarization remains a challenging research field. The extraction process depends on a set of linguistic and-or statistical features.

Linguistic-based summarizers tend to build a formal representation of the conveyed information through the text to summarize. The central intuition behind it is to employ discourse analysis techniques to model text rhetoric. For instance, the Rhetorical Structure Analysis (*RST*) can be employed to find out "*Nucleus*" fragments, which contain salient information, and "*satellite*" ones delivering additional information about the *nucleus*. In this specific case, "*Nucleus*" sentences would have higher relevancy scores, and they will be chosen to be part of the generated summary (Barzilay, 1999 and Kundi, 2014).

On the other hand, there are three variants of statistically-based summarizers; *i)* frequency-based, *ii)* feature-based, and *iii)* machine-learning-based ones. Frequency-based summarizers are built upon one of two primary hypotheses. The first one is *a)*: "cue words would be repeated many times in a given text document"; in this case, term frequency (Nenkova, 2005, 2006) is used to compute sentence relevancy scores. Inverse document frequency (a kind of probabilistic measure) (Filatova, 2004, Fung, 2006 and Galley, 2006) is used for the same reason if we assume *b)*: "essential words are more frequent in a given document than in another one.". In other words, the inverse document frequency measure estimates how relevant a word is to a given text in a set or text records. Feature-based statistical summarizers employ a bunch of indicators to compute sentence relevancy scores. Those indicators are mainly, the presence of cue headline tokens, sentence length or position, etc. (Gupta, 2010). Machine learning-based summarizers makes use of training data to learn "*relevant*" and "*non-relevant*" sentence patterns (Svore, 2007, Burges, 2005 and Hannah, 2014, Bharti, 2019).

Automatically generated summaries are regarded as straightforward, reliable, and fluid abstracts if they meet three main criteria:

- Retention: It is a measure of the extent to which the generated summary covers different topics discussed in the retrieved document (or set of documents).
- Fidelity: It is a measure of the extent to which the summary accurately reflects the author's point of view(s).
- Coherence: It is a measure of the extent to which the generated extract is semantically meaningful.

This chapter presents a *Lucene*-based document retrieval framework. It proposes two query expansion techniques: The first one uses parallel corpora while the second one bases on word embeddings to boost retrieval accuracy. Next, it adds to the proposed *DR* framework, a mono/multi-document summarization layer. A concise, reliable, and fluid extract of a given longer retrieved text document (or a set of documents) is returned as a query result instead of a crud index (or a set of indexes). The coming section describes the suggested document retrieval framework and details the proposed expansion and summarization protocols. The sixth one reports obtained results. The last section concludes this article and details ongoing and planned work.

7.5 Methodology

This section presents the suggested *Lucene*-based *DR* framework as well as the proposed query expansion techniques. Also, it describes the theoretical details of the summarization process.

7.5.1 System overview

Lucene is a robust and scalable open-source Java-based Search library. It can be easily integrated into any application to add impressive search capabilities to it. It implements the core services needed for indexing and record searching for both structured and no structured data.

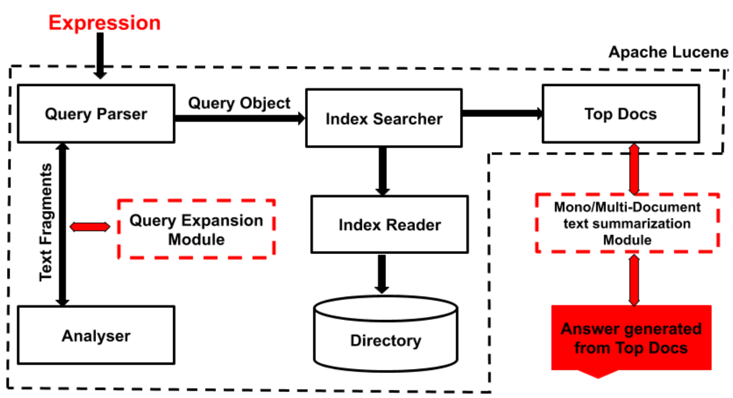

Figure 7-1: The proposed IR/DR framework

The proposed *IR/DR* framework performs the following processes shown by the above Illustration:

1. Acquiring crude contents: this first step refers to collecting the data utilized later to be queried during the retrieval phase.
2. Analyzing crude documents: It consists of merely converting each instance of the crude data to a given format that can be efficiently guessed and rendered.

3. Indexing data: It consists of mapping each document by a specific key. Next, the retrieval process will base on particular keys rather than the entire document's full content.
4. Retrieving top documents: It consists of returning indexes of top matching documents to the user query.
5. Summarizing a given retrieved document or a set of documents: It consists of returning an abstract of the top retrieved document (or a set of the top retrieved documents).

Operations *1)*, *2)*, and *3)* are generally performed off-line. Users can query the described *DR* framework; after that, all the text records are appropriately indexed. Generally, indexing documents bases either on a vectorial model (*TF-IDF*) or a probabilistic one (*BM25*). The query is converted onto a bag of words, and the index database is investigated to get the query response. Any returned reference is displayed to the user as a concise, reliable, and fluid extract of a given longer retrieved text document.

TF-IDF estimates how relevant a word *w* is to a document *d* in a corpus of text documents *D*. It is computed by multiplying two different quantities (Breitinger, 2015 and Hiemstra, 2000):

- The term frequency (*tf*): refers to the number of occurrences of w in d. Usually, the term frequency is adjusted by the d 's length or d 's most recurrent word frequency.
- The inverse document frequency (*idf*): refers to how common or rare w is in the entire corpus *D*. Being close to 0 means that the w is commonly used in *D*.

The higher the *TF-IDF* score is, the more relevant that word is in that particular document. The *TF-IDF* score for the word *w* in *d*; a document belonging to a set of documents *D* is computed as follows:

$$TFIDF(w, d, D) = tf(w, d)\,.\,idf(w, D) \quad (1)$$

where: $tf(w, d) = log(1 + freq(w, d))$ and $idf(w, D) = log(\frac{N}{count(d \in D : w \in d)})$.

BM25 (Stephen, 1994) ignores the inter-relationship between the query terms within a document. Its ranking process works as follows: Given a query *Q*, containing canonic words $q_1, ..., q_n$, the *BM25* score of a document *d* is defined as:

$$score(d, Q) = \sum_{i=1}^{n} (q_i) \frac{f(q_i, N).(k_1 + 1)}{f(q_i, N) + k_1(1 - b + b\,\frac{|N|}{avgdl})} \quad (2)$$

where $f(q_i, d)$ refers to the q_i 's term frequency in the document *d*, *N* refers to the document d 's length, and *avgdl* is the average length of all text documents. K_1 and *b* are free parameters, usually empirically fixed as $k_1 \in [1.2, 2.0]$ and $b = 0.75$.

7.5.2 Query expansion to boost retrieval accuracy

The next couple sub-sections describe the proposed two query expansion techniques. The goal is to boost the proposed document retrieval framework accuracy by making user queries more informative while preserving their integrity.

Comparable corpora-based query expansion

The first proposed technique of query expansion uses *Wikipedia* as a comparable corpus. Two slightly different variants of the same approach are described below:

- Summary-based query expansion: The Rake algorithm (Stuart, 2009) is used to extract keywords. Rake is a domain-independent keyword extraction technique. It returns a list of keywords in a text with their order of importance. The most important keyword is used as a canonic word to query Wikipedia. Following this, a one-sentence summary of the first returned Wikipedia page is generated. Next, it is concatenated to the original query.
- Content-based query expansion: the most crucial keyword returned by the RAKE algorithm is used to query Wikipedia. Next, the top returned Wikipedia page's title is concatenated to the user's query to make it more informative.

Word embeddings-based query expansion

The main idea is to expand any user query by terms having the closest embedding representation to its relevant terms. For instance, if the query contains the word "bumper," it is expanded by a set of semantically close words like "brackets," "fillers," and "parts" since "bumper," "bumper brackets," "bumper fillers" and "car parts" usually co-occur together. In this way, all text records or technical sheets related to "bumper brackets" would be considered when searching for relevant records for "bumper" even though the word "brackets" is not present in the user query. The *Gensim* implementation of *word2vec* is used to find relevant expanding terms to the user query. In this context, three distinct models were tested, namely *fasttext-wiki-news-subwords-300, glove-twitter-25, glove-twitter-200,* and *glove-wiki-gigaword-300* (Jeffrey, 2014).

7.5.3 Automatic text summarization

Automatic text summarization is used as a top layer of the proposed Lucene-based document retrieval system. It abstracts a single retrieved document (or a set of retrieved documents) to create a short, accurate, and fluent extract. In the case of multi-retrieved document summarization, all retrieved texts are concatenated and considered a single retrieved document. Mathematically, the main idea is to project the document to summarize onto a lower-dimensional space that captures the essence of concepts present in the source text. The latter space's unitary vectors are used to compute *retention-fidelity* scores, as described in our paper (Ben Ayed, 2019). The mathematical and implementation details of the proposed summarization protocol will be expanded in the coming two subsections.

Retention-Fidelity (RF) tensor construction

First, a lexicon, including all unique non-generic words, is constructed. Next, each retrieved text is segmented into n sentences. Each sentence S_i is represented by a sentence column feature vector x_i. x_i is a vector of d components. Note that d is equal to the lexicon's cardinality. Each component of x_i represents the number of occurrences of a given word of the lexicon in the text to summarize.

$$x_i = \begin{pmatrix} w_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix} \text{ and } x_i^T = (x_{i1}, x_{i2}, \dots, x_{id}) \tag{3}$$

A set of sentence feature vectors is strongly correlated if one or many components are simultaneously highly activated. For instance, if the number of occurrences of one or more tokens like "economy," "system," "private," "individuals," "businesses," "own-capital" exceeds a given threshold, it would be probable that this set of sentences are discussing the concept of "Capitalism." The goal is to project the crude sentence feature dataset from many correlated coordinates onto fewer uncorrelated ones called principal concepts while still retaining most of the original data's variability. Thus, sentence feature vectors are stacked as rows of a data matrix to construct the crude text feature matrix

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{pmatrix} \tag{4}$$

The mean sentence vector (Eq. 5) is subtracted from each sentence feature vector to remove noise and redundant information (Eq. 6). Next, the normalized text feature matrix is constructed by stacking zero-centered sentence feature vectors as its rows.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}\ , \qquad \ldots\ , \qquad \frac{1}{n}\sum_{i=1}^{n} x_{id}\right)^T \tag{5}$$

$$X = \begin{pmatrix} x_1^T - \mu^T \\ x_2^T - \mu^T \\ \vdots \\ x_n^T - \mu^T \end{pmatrix} \tag{6}$$

Next, the covariance around the mean is computed as follows:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T = \frac{1}{n-1}\ X^T X \tag{7}$$

As said bellow, the motivation is to project the crude sentence feature vectors dataset from many correlated coordinates onto fewer uncorrelated ones called principal concepts. Vectors encoding those concepts will be built sequentially in a way that maximizes their contributions to the variances of the original set of sentence feature vectors. Mathematically, the goal is to find a collection of $k \le d$ unit vectors $v_i \in \mathbb{R}^d$ (for $i \in 1, \ldots, k$) called principal concepts, such that:

The variance of the set of sentence feature vectors projected onto the v_i direction is maximized. v_i should be orthogonal to $v_1, \ldots, v_{i-1}$.

The projection of a vector $x \in \mathbb{R}^d$ onto the line determined by any v_i is simply given as the dot product $v_i^T x$. The variance of the sentence feature vector x projected onto the first principal concept $v_1$is defined as follows:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (v_1^T x_i - v_1^T \mu)^2 = \ v_1^T S v_1 \qquad (8)$$

To construct v_1, S is maximized while satisfying the $\|v_1\| = 1$ additional constraint. The Lagrange multipliers (LM) approach is used to solve this optimization problem. LM implies that $Sv_1 = \lambda_1 v_1$,aka; v_1 is an eigen concept (mathematically, it is an eigenvector of the covariance matrix S). Note that $\parallel v_1 \parallel = v_1^T v_1 = 1$, this means that the corresponding eigenvalue is equal to $v_1^T S v_1 = \ \lambda_1$. It equals the variance of the sentence feature vectors along v_1. The most important concept is coded by the eigenvector associated to the highest eigenvalue.

Next, the sentence feature vectors set is projected onto a new direction v_2 , the same way, while satisfying the $v_1 \perp v_2$ condition, then onto v_3 while satisfying $v_3 \perp v_1, v_2$, and so on. By the end of this process, the first k vectors encoding principal concepts of $X$are built. They are eigenvectors of the covariance matrix S corresponding to its k highest eigenvalues. Next, the conceptual space will be constructed such that the k most important eigen concepts will form its orthonormal basis Ξ_k:

$$\Xi_k = [v_1, v_2, \dots, v_k] \qquad (9)$$

Each normalized projected sentence onto the constructed conceptual space can be written as a linear combination of k eigen concepts. Next, the goal is to build a retention-fidelity tensor. Thus, the Euclidean distance between a given concept v_j ; $j =$

$1, \ldots, k$ and any normalized sentence $\widehat{x_i} = x_i - \mu$, projected in the conceptual space is defined and computed as follows:

$$d_i(v_j) = \|v_j - \widehat{x_i}\| \tag{10}$$

The *Retention-Fidelity* tensor provides distances between algebraic sentence feature vectors and the orthonormal conceptual space basis's unitary vectors. It is constructed such that the line order depends on the importance of a given concept, while the column order is related to the extent to which a random sentence encodes a given concept. For instance, the first line provides the *w* best sentences to encode the first most crucial concept (their normalized projected feature vectors have the smallest distances to v_1 encoding the most important concept). The second line provides the same information related to the second most important concept, and so on. Note also that the fifth sentence, for instance, is the best sentence to encode the second most crucial concept, while the sixth sentence is the last one in a window size of four sentences.

1st Eigen Concept v_1	[3 0.09]	[6 0.19]]	[12 0.32]	[4 0.66]
2nd Eigen Concept v_2	[5 0.07]	[3 0.11]	[4 0.13]	[6 0.47]
3d Eigen Concept v_3	[6 0.18]	[4 0.33]	[7 0.37]	[9 0.75]
4th Eigen Concept v_4	[5 0.22]	[6 0.24]	[7 0.29]	[2 0.65]
5th Eigen Concept v_5	[12 0.17]	[5 0.47]	[6 0.48]	[3 0.59]

: Computed **Tensor** of the first five eigen concepts with a window of 4 sentences

[*i d*] : i is a sentence index, d = distance(S_i, v_j) ; j = 1 … 5.

Figure 7-2: Th retention-Fidelity tensor

As described in the coming section, the Retention-Fidelity tensor will be used to compute a Retention-Fidelity score for each sentence.

Retention-Fidelity (RF) score computation and summary construction

First, a *Retention* score is computed for each normalized sentence being projected onto the constructed conceptual space. A given sentence having a high *Retention* sore should encode as much as possible the most important concepts expressed in the retrieved text document. In other words, it should appear as much as possible in a window of size *w* while taking into consideration the k important concepts. Mathematically, it is defined as follows:

$$R_{kw}(s) = \frac{1}{k}\sum_{i=1}^{k} \alpha_i \qquad (11)$$

$\alpha_i = 1$ if the sentence S occurs in the *i*th window. If not, it is equal to zero.

Now, an extended fidelity *(*$F_{kw}(s)$*)* score is computed for every sentence. It is a kind of averaged sum of the *retention* coefficient. The latter one is weighted according to the sentence's position in each window of size w. The central intuition is that sentences with a high F_{kw} score should encode important concepts while focusing on the most important ones. The *fidelity* score is defined as follows:

$$F\ _{kw}(s) = \frac{1}{k}\sum_{i=1}^{k} \alpha_i \left[1 + \frac{1 - \Psi_i}{w}\right] \qquad (12)$$

$\alpha_i = 1$ if a sentence s occurs in the *i*th window. If not, it is equal to zero. ψ_i is the rank of a sentence s in the *i*th window.

Next, Fuzzy logic is used to compute a unified *Retention-Fidelity* (*R-F*) score for every sentence of the retrieved document following the previously described protocol in (Ben Ayed, 2019). Highly scored sentences are extracted to generate a concise abstract as a response to the user's query.

7.6 Experiments, results, and discussion

7.6.1 The dataset

The *Trec* dataset, a news corpus of 248500 journal articles, is used for experiments. It covers many fields such as politics, economics, technology, science, etc. Crude data is preprocessed by removing stop words and applying stemming routines. The stemming technique consists of removing common endings to transform words to their root form. The most common widely used stemming algorithms are Porter, Lancaster, and Snowball. The latter is used in this project.

Precision and *recall* are commonly used to measure document retrieval effectiveness (David, 2011). *Precision* refers to the probability given that a text is retrieved; it will be relevant. *Recall* refers to the probability given that a text is relevant; it will be retrieved. In this research paper, the *TRECEVAL* program is used to evaluate the retrieval accuracy. It uses the mentioned below evaluation procedures:

- **P5:** Precision after 5 docs retrieved.
- **P100:** Precision after 1000 docs retrieved.
- **MAP:** Mean Average Precision.

Also, the *FRESA* protocol (Torres-Moreno, 2010) is used to evaluate the quality of the generated summaries.

7.6.2 Results and discussion

Query expansion effectiveness related results are reported in Tables 7-1, 7-2, 7-3, and 7-4.

Table 7-1: On the relevance of pre-processing to improve document retrieval accuracy

Data Type	**Original Data**			**Stemmed Data**		
Metric	P5	P10	Map	P5	P10	Map
Short queries	0.192	0.026	0.115	**0.196**	**0.030**	**0.148**
Long queries	0.194	0.030	0.139	**0.236**	**0.037**	**0.148**

Table 7-2: Document retrieval accuracy when using TF-IDF VS. BM25 weighting schemas

Weighting Schema	**TF-IDF**			**BM25**		
Metric	P5	P10	Map	P5	P10	Map
Short queries	0.196	0.172	0.148	**0.211**	**0.180**	**0.152**
Long queries	0.236	**0.266**	0.148	**0.242**	0.221	**0.161**

Table 7-1 compares system accuracy when using non-processed vs. pre-processed data. It confirms that pre-processing helps to reach better accuracy rates. Table 7-2 compares the obtained retrieval precisions using two different weighting schemas (*TF-IDF* and *BM25*). Note that the same pre-processing protocol was used in both scenarios. Generally, the *BM25* weighting schema outperforms the *TF-IDF* one.

Table 7-3: Obtained results when expanding queries by summary and content

Expansion Strategy	**O**			**C**			**S**		
Metric	P5	P10	Map	P5	P10	Map	P5	P10	Map
Short queries	**0.196**	**0.172**	**0.148**	0.195	0.156	0.149	0.072	0.109	0.057

Obtained results when using a comparable corpus as an expansion technique are reported in Table 7-3. Three main experiences were conducted: *O)*; precision without any expansion technique, *C)*; expanding user queries by content and *S)*; expanding user queries by summaries. Note that the same pre-processing was performed and, the same weighting schema is used for experiments *O)*, *C)*, and *S)*. Reported results in Table 7-3 show that using titles of the top returned Wikipedia pages to expand user queries provides almost the same accuracy rates as without using any query expansion technique. Using the summary of the Wikipedia top page to expand user queries messes up the retrieval precision.

Table 7-4: Obtained results when expanding queries using word embeddings

Expansion Strategy	**O**			**WE1**			**WE2**			**WE3**			**WE4**		
Metric	P5	P10	Map	P5	P10	Map	P5	P10	Map	P5	P10	Map	P5	P10	Map
Short queries	0.196	0.030	0.148	0.164	0.026	0.018	0.011	0.027	0.116	**0.204**	0.032	0.125	**0.216**	0.028	0.132

Table 7-4 reports obtained results when using word embeddings to expand user queries. Glove-twitter-25 (*WE1*), glove-twitter-200 (*WE2*), fasttext-wiki-news-subwords-300 (*WE3*), and glove-wiki-gigaword-300 (*WE4*) word2vec variants provided by The *Gensim* implementation of *word2vec* are used to perform the expansion process. Achieved results confirm that the system retrieval accuracy can be improved when considering the top 5 retried documents under a critical constraint that consists of choosing the appropriate *word2vec* model. In the bellow example, *WE3*, which is trained using a collection of news articles, and *WE4*, which is trained using a massive corpus of textual data, helped ameliorate retrieval accuracy. It was not the case when using the inappropriate *word2vec* model to this specific context.

Table 7-5: Obtained Fresa scores for mono/multi document summarization with window sizes w = 2,4 and 6

Window size (*w*)	**Mono-Document Summarization**	**Multi-Document summarization (*D* = 2)**	**Multi-Document summarization (*D* = 5)**	**Multi-Document summarization (*D* = 10)**
w = 2	0.642	0.590	0.398	0.136
w = 4	**0.748**	**0.711**	0.473	0.213
w = 6	0.219	0.340	**0.591**	0.311

The *FREZA* evaluation protocol (Torres-Moreno, 2010) was used to evaluate the quality of the generated abstracts (Table 7-5). The best results are obtained when we summarize the first retrieved document. For multi-document summarization, the best results are obtained when we summarize fewer retrieved documents (D =2) with a window size $w = 4$. If we want to summarize more than two retrieved documents (the top five ones, for instance) while approximately preserving the same quality of the returned result, we should consider a bigger window size ($w = 6$). Generally, summarizing many documents ($D = 10$) deteriorates the quality of the final query response.

7.7 Conclusion and future work

This paper presented a *Lucene* based document retrieval framework. Comparing two different weighting schemas (*TF-IDF* and *BM25*) shows that the *BM25* probabilistic model outperforms the vectorial one (*TF-IDF*). Additionally, led query expansion experiments show that using word embeddings enhances the overall document retrieval precision. It is not the case when using a comparable corpus. The proposed framework can be enhanced by implementing an interactive query expansion approach: The obtained result using the proposed comparable corpus-based expansion approach depends on the efficiency of the *Rake* keyword extractor algorithm. The central intuition is to involve users in the query expansion process. Users have to approve the returned extracted keywords. Another hybrid technique of query expansion may be explored: Once the user validates keywords, a word embeddings expansion will be performed. The latter technique will ensure using only appropriate keywords and retrieving relevant records that do not necessarily contain terms used in the user query.

Mono-document and multi-document summarization of the few top retrieved references achieved excellent coverage and fidelity levels, fundamental criteria of useful summaries. Note that the coherence of the generated query response is out of the scope of this paper. Applying a discourse analysis technique like the Rhetorical structure theory (*RST*) establishes a formal representation of the retrieved document's knowledge. It helps to generate more coherent abstracts (Mann, 1988). Achieving a fully coherent abstract of mono-document summaries is straightforward. Employing the rhetorical structure theory technique, as mentioned above, can quickly achieve it. For multi-document summarization, local coherence can be achieved, and the main challenge would be to achieve a global coherence. Finally, the proposed summarization protocol can be improved to create a more professional human-like abstract of the top retrieved documents; Rephrasing techniques can be performed on extracted text segments to generate original sentences instead of merely extracting the most salient ones.

RÉFÉRENCES

[Abdel-Fattah 2009] Mohamed Abdel-Fattah and Fuji Ren. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. In Computer Speech & Language, pages 23, pages 126-144, 2009.

[Ahmed 2013] Ibrahim Ahmed, and Tarek Elghazaly. Improve the automatic summarization of Arabic text depending on Rhetorical Structure Theory. In The 12th Mexican International Conference on Artificial Intelligence (MICAI). IEEE, pages 223-227, 2013.

[Abouenour 2013] Lahsen Abouenour, Karim Bouzoubaa, Paolo Rosso. On the evaluation and improvement of Arabic WordNet coverage and usability. In Language Resources and Evaluation, volume 47, issue 3, pages 891-917, 2013.

[Abualigah 2020] Laith Abualigah, Bashabsheh Mohammad, Alabool, Hamzeh and Shehab Mohammad. Text Summarization: A Brief Review. In Recent Advances in NLP: The Case of Arabic Language, volume 874, pages 1-15. 2020.

[Albanese 2006] M. Albanese, M. Fayzullin, A. Picariello et V.S. Subrahmanian. The priority curve algorithm for video summarization. Information Systems, vol. 31, no. 7, pages 679–695, November 2006.

[Albrecht 1995] Jason E. Albrecht and Jerome L. Myers. Role of context in accessing distant information during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, volume 21, pages 1459-1468, 1995.

[Allahyari 2017] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, Krys Kochut. Text Summarization Techniques: A Brief Survey. Journal of Computation and Language, volume 1, 2017.

[Andhale 2016] Narendra Andhale and L.A. Bewoor. An overview of Text Summarization techniques. In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA), pages 1-7, 2016.

[Asghar 2014] Muhammad Z. Asghar. Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet. Life Science Journal, volume 11, 2014.

[Al-Abdallah 2017] Raed Z. Al-Abdallah and Ahmad T. Al-Taani. Arabic Single-Document Text Summarization Using Particle Swarm Optimization Algorithm. Journal of Procedia Computer Science, volume 117, pages 30-37, 2017.

[Al-Qassem 2017] Lamees M. Al-Qassem, Di Wang, Zaid Al Mahmoud, Hassan Barada, Ahmad Al-Rubaie and Nawaf I. Almoosa. Automatic Arabic Summarization: A survey of methodologies and systems. In Procedia Computer Science journal, volume 117, pages 10-18, 2017.

[Al-Saleh 2016] Asma B. Al-Saleh and Mohamed El Bachir Menai. Automatic Arabic text summarization: a survey. In Artificial Intelligence Review. Volume 45, pages 203-234, 2016.

[Al-Sanie 2005] Waleed Al-Sanie. Towards an infrastructure for Arabic ext summarization using rhetorical structure theory. Master's Thesis. King Saud University, Riyadh, Saudi Arabia. 2005.

[Al-Taani 2014] Ahmad Al-Taani and Maha Al-Omour. An Extractive Graph-based Arabic Text Summarization Approach. In the International Arab conference on information technology (ACIT); pages 158-163, 2014.

[Azmi 2012] Aqil Azmi and Suha S. Althunayyan. A text summarizer for Arabic. Journal of Computer Speech and Language, volume 26, issue 4, pages 260-273, 2012.

[Asher 2005] Nicholas Asher and Alex Lascarides. Logics of conversation. In Cambridge University Press, 2005.

[Barzilay 1997] Regina Barzilay et Michael Elhadad. Using lexical chains for text summarization. In Proceedings of the ACLWorkshop on Intelligent Scalable Text Summarization, pages 10–17, 1997.

[Bär 2015] Daniel Bär, Torsten Zesch, and Iryna Gurevych. Composing Measures for Computing Text Similarity. Technical Report TUD-CS-2015-0017, TU Darmstadt, Allemagne, 2015.

[Belguith 2014] Lamia H. Belguith, Mariem Ellouze, Mohamed H. Maaloul, Maher Jaoua, Fatma K. Jaoua and Philippe Blache. Automatic summarization. Natural Language Processing of Semitic Languages. Springer Berlin Heidelberg, pages 371-408, 2014.

[Belkebir 2015] Riadh Belkebir and Ahmed Guessoum. A supervised approach to Arabic text summarization using adaboost. In New contributions in information systems and technologies, Springer International Publishing, volume 353. pages 1-23, 2015.

[Ben Ayed 2019] Alaidine Ben Ayed, Ismaïl Biskri and Jean-Guy Meunier: Automatic Text Summarization: A New Hybrid Model Based on Vector Space Modelling, Fuzzy Logic and Rhetorical Structure Analysis. ICCCI, volume 2, pages 26-34. 2019

[Bharti 2019] Sharma Bharti, Nitika Katyal, Vishant Kumar, Shivani, Amit Lathwal, and Khanna, Ashish. Automatic Text Summarization Using Fuzzy Extraction. In International Conference on Innovative Computing and Communications, Volume 1, pages 395–405. 2019.

[BLACK 2006] William BLACK, Sabri Elkateb, Piek Vossen, Horacio Rodriguez, Musa Alkhalifa, Adam Pease and Christiane Fellbaum. Introducing the Arabic WordNet Projec. In Proceedings of LREC, 2006.

[Boudin 2009] Florian Boudin and Juan-Manuel Torres-Moreno. Résumé automatique multi-document et indépendance de la langue : une première évaluation en français. In Traitement Automatique des Langues Naturelles (TALN), 2009.

[Burges 2005] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton and Greg Hullender. Learning to rank using gradient descent. In Proceedings of the 22nd international conference on Machine learning (ICML '05), pages 89–96, 2005.

[Cabrera-Diego 2018] Luis Adrián Cabrera-Diego and Juan-Manuel Torres-Moreno. Summtriver: A new trivergent model to evaluate summaries automatically without human references. Data & Knowledge Engineering, 113 :184–197, 2018.

[Cabrera-Diego 2016] Luis Adrián Cabrera-Diego, Juan-Manuel Torres-Moreno, and Barthélémy Durette. Evaluating multiple summaries without human models: A first experiment with a trivergent model. In International conference on applications of natural language to information systems, pages 91–101. Springer, 2016.

[Carson 2002] Chad Carson, Serge Belongie, Hayit Greenspan and Jitendra Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pages 1026–1038, August 2002.

[Chakraborty 2019] Roshni Chakraborty, Maitry Bhavsar, Sourav Kumar Dandapat and Joydeep Chandra.Tweet Summarization of News Articles: An Objective Ordering-Based Perspective. In IEEE Transactions on Computational Social Systems, vol. 6, no. 4, pages 761-777. 2019.

[Chen 2021] Yi-Syuan Chen and Hong-Han Shuai. Meta-Transfer Learning for Low-Resource Abstractive Summarization. arXiv. 2021.

[Cheung 2002] Jackie Chi Kit Cheung, Hoifung Poon, and Lucy Vanderwende. Probabilistic frame induction. arXiv preprint arXiv :1302.4813, 2013.

[Chin-Yew 2004] Lin Chin-Yew. ROUGE: a Package for Automatic Evaluation of Summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, 2004.

[Conroy 2010] John M. Conroy, Judith D Schlesinger, Peter A Rankel, and Dianne P O'Leary. Guiding classy toward more responsive summaries. In TAC, 2010.

[Conroy 2008] John M. Conroy and Hoa Trang Dang. Mind the gap: Dangers of divorcing evaluations of summary content from linguistic quality. In: Proceedings of the 22Nd International Conference on Computational Linguistics, volume 1, issue8. USA, PA, USA, pages 145-152. 2008.

[Cox 2018] D.R. Cox, Christiana Kartsonaki and Ruth H.Keogh. Big data: Some statistical issues. Journal of Statistics & Probability Letters, Volume 136, pages 111-115. 2018.

[Dutta 2019] Soumi Dutta, Vibhash Chandra, Kanav Mehra, Sujata Ghatak, Asit K. Das and Saptarshi Ghosh. Summarizing Microblogs During Emergency Events: A Comparison of Extractive Summarization Algorithms. In: Abraham A., Dutta P., Mandal J., Bhattacharya A., Dutta S. (eds) Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, volume 813. Springer, Singapore. 2019.

[Edmundson 1969] Harold P. Edmundson. New Methods in Automatic Extracting. Journal of the Association for Computing Machinery, vol. 16, no. 2, pages 264–285, April 1969

[El-Haj 2013] Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. Experimenting with automatic text summarisation for Arabic. Journal of Language and Technology, Springer Berlin Heidelberg, pages 490-499. 2013.

[EL-Haj 2010] Mahmoud EL-Haj and Udo Kruschwitz. Using Mechanical Turk to create a corpus of Arabic summaries. In Language Resources (LRs) and Human Language Technologies (HLT) for Semitic Languages workshop held in conjunction with the 7th International Language Resources and Evaluation Conference (LREC 2010), 2010.

[EL-Haj 2011] Mahmoud EL-Haj and Udo Kruschwitz. Exploring clustering for multi-document Arabic summarisation. In Asia Information Retrieval Symposium, Springer Berlin Heidelberg, pages 550-561, 2011.

[El-Khair 2017] Abu El-Khair, Ibrahim. Effects of stop words elimination for Arabic information retrieval: a comparative study. International Journal of Computing & Information Sciences, volume 4, issue 3, pages 119-133, 2017.

[El-Kassas 2021] Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea and Hoda K. Mohamed. Automatic text summarization: A comprehensive survey. In Expert Systems with Applications, volume 165,2021.

[Ekin 2003] A. Ekin, A. M. Tekalp et R. Mehrotra. Automatic soccer video analysis and summarization. Image Processing, IEEE Transactions on, vol. 12, no. 7, pages 796–807, July 2003.

[Erkan 2004] Gunes Erkan and Dragomir R. Radev. LexRank: Graph-Based Lexical Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research, volume 22, pages 457-479. 2004.

[Farzindar 2004] Atefeh Farzindar, Guy Lapalme et Jean-Pierre Desclés. Résumé de textes juridiques par identification de leur structure thématique. Traitement automatique de la langue (TAL), vol. 45, no. 1, pages 39–64, jan 2004.

[Fei-Fei 2003] Li Fei-Fei, Rob Fergus and Pietro Perona. A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories. In Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2, ICCV ’03, pages 1134–, Washington, DC, USA, 2003. IEEE Computer Society.

[Fejer 2014] Hamzah N. Fejer and Nazlia Omar. Automatic Arabic text summarization using clustering and keyphrase extraction. In Proceedings of the 6th Information Technology and Multimedia (ICIMU), pages 293-298, Putrajaya, Malaysia. 2014.

[Filatova 2004] Elena Filatova, Vasileios Hatzivassiloglou. A formal model for information selection in multi-sentence text extraction. In: COLING 2004; The 20th International Conference on Computational Linguistics (ACL), pages 397-403, 2004.

[Filippova 2010] Katja Filippova. Dependency Graph-Based Sentence Fusion and Compression. PhD thesis, TU Darmstadt, Allemagne, 2010.

[Frank 2003] Stefan Frank, Mathieu Koppen, Leo Noordman and Wietske Vonk. Modeling knowledge-based inferences in story comprehension, Cognitive Science, volume 27, pages 875-910, 2003.

[Fung 2006] Pascale Fung and Grace Ngai. One story, one flow: Hidden Markov story models for multilingual multidocument summarization. In: The ACM Trans. Speech Lang. Process, pages 1-16, 2006.

[Galley 2006] Michel Galley. A skip-chain conditional random field for ranking meeting utterances by importance. In: The Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pages 364-372, 2006.

[Gambhir 2017] Mahak Gambhir and Vishal Gupta. Recent automatic text summarization techniques: a survey. The Artificial Intelligence Review, volume 47, issue 1, pages 1-66, 2017.

[Généreux 2009] Michel Généreux and Aurélien Bossard. Résumé automatique de textes d'opinions. In TALN 2009, 2009.

[Genest 2012] Pierre-Etienne Genest and Guy Lapalme. Fully Abstractive Approach to Guided Summarization. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 354–358, 2012.

[Genest 2011] Pierre-Etienne Genest and Guy Lapalme. Framework for Abstractive Summarization using Text-to-Text Generation. In Proceedings of the Workshop on Monolingual Text-To-Text Generation, pages 64–73, Portland, Oregon, June 2011.

[Gernsbacher 1995] Morton Ann Gernsbacher. The Structure Building Framework: What it is, what it might also be, and why. In B. K. Britton and A. C. Graesser, (Eds.), Models of text understanding, pages 289-311, 1995.

[Giang 2013] Binh Giang, Tuan A. Tran, Nam K. Tran and Mohammad Tran. Leverage Learning to rank in an optimization framework for timeline summarization. In: TAIA workshop SIGIR. (2013)

[Golden 1994] Richard M. Golden, David E. Rumelhart, Joseph Strickland and Alice Ting. Markov random fields for text comprehension. In: Levine D.S. and Aparicio M. (Eds.), Neural networks for knowledge representation and inference, pages 283-309, 1994.

[Golden 1993] Richard M. Golden and David E. Rumelhart. A parallel distributed processing model of story comprehension and recall. Discourse Processes, volume 16, pages 203-237, 1993.

[Gonz'alez-Gallardo 2020] Carlos-Emiliano Gonz'alez-Gallardo, Romain Deveaud, E. SanJuan, Juan-Manuel Torres. Audio Summarization with Audio Features and Probability Distribution Divergence. Gonz'alez-Gallardo, C., Deveaud, R., SanJuan, E., & Torres, J. (2020). Audio Summarization with Audio Features and Probability Distribution Divergence. ArXiv, abs/2001.07098. 2020.

[Grabowski 1992] Joachim Grabowski. Expository text and propositional text processing. In: Hout-Wolters, Bernadette and Schnotz, Wolfgang (Eds.), Text Comprehension and Learning from Text. Amsterdam: Swets and Zeitlinger, pages 19-33,1992.

[Guellil 2019] Imane Guellil, Houda Saadane, Faical Azouaou, Billel Gueni and Damien Nouvel. rabic natural language processing: An overview. Journal of King Saud University - Computer and Information Sciences, volume 32, issue 10, pages 1227--1228, 2019.

[Gupta 2010] Vishal Gupta, Gurpreet Singh Lehal. A survey of text summarization extractive techniques. Journal of Emerging Technologies in Web Intelligence, volume 2, issue 3, pages 258-268, 2010.

[Gupta 2012] Virenda K. Gupta and Tanveer J. Siddiqui. Multi-document summarization using sentence clustering, In the 4th International Conference on Intelligent Human Computer Interaction (IHCI), pages 1-5, 2012.

[Haboush 2012] Ahmad Haboush, Maryam Al-Zoubi, Ahmad Momani and Motassem Tarazi. Arabic text summarization model using clustering techniques. Journal of Word of Computer Science and Information Technology (WCSIT), volume 2, issue 3, pages 62-67, 2012.

[Hannah 2014] Esther Hannah and Saswati Mukherjee. A classification-based summarisation model for summarising text documents. nternational Journal of Information and Communication Technology, volume 6, issue 3, pages 292-308, 2014.

[Hong 2014] Hong Kai Hong, John Conroy, Benoit Favre, Alex Kulesza, Hui Lin, and Ani Nenkova. A Repository of State of the Art and Competitive Baseline Summaries for Generic News Summarization. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 1608–1616, Reykjavik, Iceland, May 2014. ELRA. ISBN 978-2-9517408-8-4.

[Saggion 2002] Horacio Saggion and Guy Lapalme. Generating Indicative-Informative Summaries with SumUM. Computational Linguistics, 28(4) :497–526, jan 2002.

[Hori 2001] Chiori Hori et Sadaoki Furui. Advances in automatic speech summarization. In Proceedings of the 7th European Conference on Speech Communication and Technology, pages 1771–1774, 2001.

[Hovy 2006] Eduard Hovy, Chin-Yew Lin, Liang Zhou et Junichi Fukumoto. Automated Summarization Evaluation with Basic Elements. In Proceedings of the Fifth Conference onLanguage Resources and Evaluation (LREC), pages 899–902, 2006.

[Hovy 2005] Eduard Hovy. Text Summarization. In: The Oxford Handbook of Computational Linguistics. 2005.

[Hovy 1998] Eduard Hovy and Chin-Yew Lin. Automated text summarization and the SUMMARIST system. In Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998, pages 197–214. Association for Computational Linguistics, 1998.

[Ibrahim 2013] Ahmed Ibrahim, Tarek Elghazaly and Mervat Gheith. A Novel Arabic text summarization model based on rhetorical structure theory and vector space model. In International Journal of Computational Linguistics and Natural Language Processing, volume 1, 2013.

[Imam 2013] Ibrahim Imam, Nihal Nounou, Alaa Hamouda and Hebat Allah A. Khalek. An ontology-based summarization system for Arabic documents (ossad). In computer application (IJCA), volume 74, issue 17, pages 38-43, 2013.

[Inoue 2004] Akira Inoue, Takayoshi Mikami et Yoichi Yamashita. Improvement of Speech Summarization Using Prosodic Information. In Proc. Speech Prosody, 2004.

[Jaradat 2016] Yazan A. Jaradat and Ahmad T. Al-Taani. Hybrid-based Arabic single-document text summarization approach using genEtic algorithm. In The 7th

International Conference on in Information and Communication Systems (ICICS2016), Irbid, Jordan, 2016.

[Joshi 2018] Monika Joshi, Hui Wang and Sally McClean. Dense Semantic Graph and Its Application in Single Document Summarisation. In: Lai C., Giuliani A., Semeraro G. (eds) Emerging Ideas on Information Filtering and Retrieval. Studies in Computational Intelligence, volume 746. Springer, 2018.

[Jaidka 2013] Kokil Jaidka, Christopher Khoo, and Jin-Cheon Na. Deconstructing human literature reviews – a framework for multi-document summarization. In Proceedings of the 14th European Workshop on Natural Language Generation, pages 125–135, Sofia, Bulgaria, August 2013.

[Kazantseva 2010] Anna Kazantseva and Stan Szpakowicz, Summarizing short stories, Computational Linguistics, volume 36, Issue 1, pages 71-109. 2010.

[Kazantseva 2006] Anna Kazantseva. An approach to summarizing short stories. In Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, EACL '06, pages 55–62, Stroudsburg, PA, USA, 2006.

[Khan 2014] Atif Khan and Naomie Salim. A Review on Abstractive Summarization Methods. Journal of Theoretical and Applied Information Tech nology, 59(1) :64–72, 2014.

[Keskes 2015] Iskandar Keskes. Discourse analysis of arabic documents and application to automatic summarization. Doctoral Thesis. University of Toulouse III-Paul Sabatier. 2015.

[Kintsch 2001] Walter Kintsch. Predication. Cognitive Science, volume 25, pages 173-202, 2001.

[Kintsch 1991] Walter Kintsch and David M. Welsch. The construction-integration model: A framework for studying memory for text. In: W. E. Hockley and S. Lewandowsky (Eds.), Relating theory and data: Essays on human memory in honor of Bennet B. Murdock. pages367-385, 1991.

[Kintsch 1988] Walter Kintsch. The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. Psychological Review, volume 95, issue 2, pages 163-182. 1988.

[Kintsch 1978] Walter Kintsch and Teun A. van Dijk. Toward a model of text comprehension and production. Psychological Review, volume 85, pages 363-394, 1978.

[Kumar 2016] Yogan J. Kumar, Ong S. Goh, Halizah Basiron, Ngo H. Choon and Puspalata C. Suppiah. A Review on Automatic Text Summarization Approaches. Journal of Computer Science, volume 2, issue 4, pages 178-190, 2016.

[Kupiec 1995] Julian Kupiec, Jan Pedersen et Francine Chen. A trainable document summarizer. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '95, pages 68–73, New York, NY, USA, 1995. ACM.

[Lagrini 2017] Samira Lagrini, Mohammed Redjimi and Nabiha Azizi. Automatic Arabic Text Summarization Approaches. In Automatic Arabic Text Summarization Approaches, volume 164, pages 31-37. 2017

[Landauer 1998] Thomas K Landauer, Peter W. Foltz and Darrell Laham. Introduction to Latent Semantic Analysis, Discourse Processes, volume 25, pages 259-284, 1998.

[Langston 1999] Mark C. Langston, Tom Trabasso and Joseph P. Magliano. A connectionist model of narrative comprehension. In A. Ram and K. Moorman (Eds.), Understanding language understanding: Computational models of reading, Cambridge, MA: MIT Press, pages 181-22,1999.

[Larkey 2002] Leah S. Larkey, Lisa A. Ballesteros and Margaret E. Connell. Improving stemming for Arabic information retrieval: light stemming and cooccurrence analysis. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pages 275-282. 2002.

[Liang 2017] Hai Liang and King-wa Fu. Information overload, similarity, and redundancy: unsubscribing information sources on twitter. Journal of Computer-Mediated Communiccation, volume 22, issue 1, pages 1-17. 2017.

[Lee 2017] Sun K. Lee, Nathan J. Lindsey and Kyun SooKim. The effects of news consumption via social media and news information overload on perceptions of journalistic norms and practices. Comput. Hum. Behav., (75), pages pages 254-263. 2017.

[Lin 1997] Chin-Yew Lin and Eduard Hovy. Identifying Topics by Position. In Proceedings of the Fifth Conference on Applied Natural Language Processing, pages 283– 290, Washington, DC, USA, March 1997.

[Lin 2004] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, pages 74–81, Barcelona, Spain, July 2004.

[Lin 2003] Chin-Yew Lin and Eduard Hovy. Automatic Evaluation of Summaries Using N-gram Cooccurrence Statistics. In Proceedings of Language Technology Conference. HLT-NAACL. 2003.

[Liu 2009] Shixia Liu, Michelle X. Zhou, Shimei Pan, Weihong Qian, Weijia Cai et Xiaoxiao Lian. Interactive, topic-based visual text summarization and analysis. In Proceedings of the 18th ACM conference on Information and knowledge management, CIKM ’09, pages 543–552, New York, NY, USA, 2009. ACM.

[Lloret 2012] Elena Lloret and Manuel Palomar. Text summarisation in progress: a literature review. Artificial Intelligence Review, volume 37, issue 1, pages 1-41. 2012.

[Louis 2008] Annie Louis and Ani Nenkova. Automatic summary evaluation without human models. In Proceedings of the First Text Analysis Conference (TAC 2008), 2008.

[Louis 2013] Annie Louis and Ani Nenkova. Automatically assessing machine summary content without a gold standard. Computational. Linguistics, volume 39, issue 2, pages 267–300, 2013.

[Lovinger 2019] Justin Lovinger, Iren Valova, and Chad Clough. Gist: general integrated summarization of text and reviews. J. Soft Computing, volume 23, issue 5, pages 1589–1601. 2019.

[Li 2019] Jun Li, Guimin Huang, Chunli Fan, Zhenglin Sun, and Hongtao Zhu. Key Word Extraction for Short Text Via Word2vec, Doc2vec, and Textrank. Turkish Journal of Electrical Engineering and Computer Science, colume 3, issue 27, pages 1794–1805. 2019.

[Luhn 1958] H. P. Luhn. The automatic creation of literature abstracts. IBM Journal of Research Development, 2(2) :159–165, 1958.

[Maaloul 2013] Mohamed H. Maaloul. Approche Hybride pour le Résumé Automatique de Textes Application à la langue arabe. Doctoral Thesis, Université de Provence-Aix-Marseille I, France. 2013.

[Madnani 2010] Nitin Madnani and Bonnie J. Dorr. Generating phrasal and sentential paraphrases: A survey of data-driven methods. Computational Linguistics, 36(3): 341–387, 2010.

[Mani 2002] Iderjeet Mani, Gary Klein, David House, Lynette Hiershman , ThereseE Fir;min and Beth Sundheim. Summac: a text summarization evaluation. Natural Language Engineering, volume 8, issue1, pages 43-68. 2002.

[Mani 2001] Inderjeet Mani. Automatic summarization. John Benjamins Publishing. (2001)

[Mani 1999] Inderjeet Mani. Advances in automatic text summarization. MIT Press, Cambridge,MA, USA, 1999.

[Mann 1988] William C. Mann and Sandra A. Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. In Text Journal, volume 8, pages 243-281, 1988.

[Marcu 1998] Daniel Marcu. Improving Summarization Through Rhetorical Parsing Tuning. In: The Sixth Workshop on Very Large Corpora, pages 206–215, 1998.

[Mathkour 2009] MaHassan I. Mathkour. A Novel rhetorical structure approach for classifing Arabic security documents. Journal of Computer Theory and engineering, volume 1, issue 3, pqges 179-2011. 2009.

[McKeown 1995] Kathleen McKeown and Dragomir R. Radev. Generating summaries of multiple news articles. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR ’95, pages 74–82, New York, NY, USA, 1995. ACM.

[McKeown 1999] Kathleen R. McKeown, Judith L. Klavans, Vasileios Hatzivassiloglou, Regina Barzilay, and Eleazar Eskin. Towards Multidocument Summarization by Reformulation: Progress and Prospects. In Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative Applications of Artificial Intelligence, AAAI ’99/IAAI ’99, pages 453–460, Menlo Park, CA, USA, 1999.

[Mihalcea 2004a] Rada Mihalcea. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In Proceedings of the ACL Interactive Poster and Demonstration Sessions, 2004.

[Mihalcea 2004b] Rada Mihalcea and Paul Tarau. TextRank: Bringing Order into Texts. In: Proceedings of Empirical Methods for Natural Language Processing, pages 404-411, 2004.

[Miller 1995] George A. Miller. WordNet: a lexical database for English. Commun. ACM, vol. 38, no. 11, pages 39–41, November 1995.

[Minel 2003] Jean-Luc Minel. (2003). Filtrage sémantique. Du résumé à la fouille de textes. Paris, Hermès, 2003.

[Mirani 2017] Tarun B. Mirani and Sreela Sasi. Two-level text summarization from online news sources with sentiment analysis, International Conference on Networks & Advances in Computational Technologies (NetACT), pages 19-24. 2017.

[Mitkov 1993] Ruslan Mitkov. Automatic Abstracting in a Limited Domain. pages 204-210, 1993.

[Mustafa 2012] Aqil M. Mustafa and Suha S Al-Thanyyan. A Text summarizer for Arabic. In Journal of Computer Speech and Language, volume 26, 260-273, 2012.

[Myers 1998] Jerome L. Myers and Edward J O'Brien. Accessing the discourse representation during reading. Journal of Discourse Processes, volume 26, pages 131-157, 1998.

[Myers 1987] Jerome L. Myers, Makiko Shinjo and Susan A. Duffy. Degree of causal relatedness and memory. Journal of Memory and Language, volume 26, pages 453-465, 1987.

[Nenkova 2006] Ani Nenkova, Lucy Vanderwende and Kathleen McKeown. A compositional context sensitive multi-document summarizer: Exploring the factors that influence summarization. In: the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 573-580, 2006.

[Nenkova 2005] Ani Nenkova and Lucy Vanderwende. The Impact of Frequency on Summarization. Microsoft Research, 2005.

[Nenkova 2004] Ani Nenkova and Rebecca J. Passonneau. Evaluating content selection in summarization: The pyramid method. In HLT-NAACL, pages 145–152, 2004.

[Neto 2003] Joel Larocca Neto, Alex A. Freitas, and Celso A. A. Kaestner. Automatic text summarization using a machine learning approach. In Advances in Artificial Intelligence, pages 205–215. Springer Berlin Heidelberg, 2003.

[Noriega 2018] Alan R. Noriega, Reyes Juárez-Ramírez, Samantha Jiménez and Sergio Inzunza. Ashur: Evaluation of the relation summary-content without human reference using rouge. Journal of Computing and Informatics, volume 37, pages 509-532, 2018.

[Oufaida 2014] Houda Oufaida, Omar Nouali and Philippe Blache. Minimum redundancy and maximum relevance for single and multi-document Arabic text summarization. In Journal of King Saud University Computer and Information Sciences, volume 26, issue 4, pages 450-461. 2014.

[Ono 1994] Kenji Ono, Kazuo Sumita et Seiji Miike. Abstract generation based on rhetorical structure extraction. In Proceedings of the 15th conference on Computational linguistics, Volume 1, COLING ’94, pages 344–348, Stroudsburg, PA, USA, 1994. Association for Computational Linguistics.

[Over 2007] Paul D. Over, Hoa T. Dang, Donna K. Harman. DUC in context. IPM, volume 43, issue 6, pages 1506-1520. 2007.

[Paice 1993] Chris D. Paice and Paul A. Jones. The identification of important concepts in highly structured technical papers. In Research and Development in Information Retrieval, pages 69–78, 1993. 8, 61.

[Passonneau 2013] Rebecca J. Passonneau, Emily Chen, Weiwei Guo, and Dolores Perin. Automated pyramid scoring of summaries using distributional semantics. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 143–147, Sofia, Bulgaria, August 2013.

[Paulus 2017] Romain Paulus, Caiming Xiong and Richard Socher. A Deep Reinforced Model for Abstractive Summarization, arXiv:1705.04304. 2017.

[Radev 2004] Dragomir R. Radev, Hongyan Jing, Malgorzata Stys, and Daniel Tam. Centroid-based summarization of multiple documents. Information Processing and

Management, 40(6) :919–938, November 2004. ISSN 0306-4573. doi : 10.1016/j.ipm. 2003.10.006.

[Radev 2003] Dragomir R. Radev, Simone Teufel, Horacio Saggion,Wai Lam, John Blitzer, Hong Qi, Arda Çelebi, Danyu Liu, and Elliott Drabek. Evaluation challenges in large scale document summarization. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pages 375–382, Sapporo, Japan, July 2003.

[Radev 2002] Dragomir R. Radev, Eduard Hovy and Kathleen McKeown. Introduction to the special issue on summarization. Computational Linguistics, pp. 399-408, 2002.

[Raffel 2019] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-TextTransformer. arXiv:1910.10683. 2019.

[Rahman 2010] Nazreena Rahman and Bhogeswar Borah. Improvement of query-based text summarization using word sense disambiguation. J. Complex Intelligent systems, volume 6, pages 75–85. 2020.

[Ramu 2019] Nikhil A. Ramu, Mohana S. Bandarupalli, Manoj S. Surya-Nekkanti and Gowtham Ramesh. Summarization of Research Publications Using Automatic Extraction. In: Hemanth D., Shakya S., Baig Z. (eds) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies, volume 38. Springer, Cham. 2019.

[Raza 2019] Muhammad Raza, Mokhtar, Rahmah and Ahmad, Noraziah. A survey of statistical approaches for query expansion. J. Knowledge and Information Systems. volume 61, 2019.

[Reeve 2007] Lawrence H. Reeve, Hyoil Han et Ari D. Brooks. The use of domain-specific concepts in biomedical text summarization. In Information Processing and Management 43, volume 43, pages 1765–1776. Elsevier Ltd., 2007.

[Roetzel 20019] Peter G. Roetzel. Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. Journal of Business Research, volume 12, pages479–522, 2019.

[Rojas 2021] Jonathan Rojas, Yulia Ledeneva and René A. G. Hernández. Evaluation of text summaries without human references based on the linear optimization of content metrics using a genetic algorithm. In: Expert Systems with Applications, Volume 167. 2021.

[Saggion 2013] Horacio Saggion and Thierry Poibeau. Automatic text summarization: Past, present and future. In: Poibeau T, SaggionH, Piskorski J, Yangarber R (eds) Multi-source, multilingual information extraction and summarization, theory and applications of natural language processing. Springer, pp. 3-21. 2013.

[Saggion 2010] Horacio Saggion, Juan-Manuel Torres-Moreno, Iria da Cunha, Eric SanJuan, and Patricia Velazquez Morales. Multilingual summarization evaluation without human models. In 23rd COLING International Conference on Computational Linguistics (Posters), pages 1059–1067, 2010.

[Sahni 2018] Aashka Sahni and Sushila Palwe. Topic Modeling on Online News Extraction. J. Intelligent Computing and Information and Communication, volume 673, pages 611-622, 2018.

[Sahoo 2016] Deepak Sahoo, Rakesh Balabantaray, Mridumoni Phukon and Saibali Saikia. Aspect based multi-document summarization. International Conference on Computing, Communication and Automation (ICCCA), pages 873-877. 2016.

[Salton 1997] Gerard Salton, Amit Singhal, Mandar Mitra and Chris Buckley, Automatic text structuring and summarization, Information Processing & Management, Volume 33, Issue 2, 1997, Pages 193-207, ISSN 0306-4573.

[Salton 1996] Gerard Salton, Amit Singhal, Chris Buckley, and Mandar Mitra. Automatic text decomposition using text segments and text themes. In Proceedings of the the seventh ACM conference on Hypertext, pages 53–65. ACM, 1996.

[Sammut 2010] Claude Sammut and Geoffrey I. Webb. Encyclopedia of Machine Learning. Springer US, 2010.

[Schiffman 200] Barry Schiffman, Ani Nenkova, and Kathleen McKeown. Experiments in multidocument summarizatbion. In Proceedings of the Second International Conference on Human Language Technology Research, HLT ’02, pages 52–58, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[Schmitt 2017] Josephine B. Schmitt, Christina A. Debbelt and Frank M. Schneider. Too much information? Predictors of information overload in the context of online news exposure. Journal of Information Communication and Society, volume 21, issue 8, pages 1151–1167, 2017.

[See 2017] Abigail See, Peter J. Liu and Christopher D. Manning. Get to The Point: Summarization with Pointer-Generator Networks. arXiv, 2017.

[Sethi 2017] Prakhar Sethi, Sameer Sonawane and Saumitra Khanwalke. Automatic text summarization of news articles. International Conference on Big Data, IoT and Data Science (BID), pages 23-29, 2017.

[Smith 2018] Stephen M. Smith and Thomas E. Nichols, Statistical Challenges in "Big Data" Human Neuroimaging, Journal of NEUROVIEW, VOLUME 97, ISSUE 2, pages 263-268. 2018.

[Sobh 2009] Ibrahim Sobh, Nevin Darwish and Magda Fayek. An optimized dual classification system for Arabic extractive generic text summarization. Doctoral Thesis. Faculty of engineering,Cairo university, Giza, Egypt. 2009.

[Spärck Jones 2007] Karen Spärck Jones, Automatic summarising: The state of the art. Inf. Process. Manage, 43(6) :1449–1481, 2007.

[St. John 1992a] Mark F. St. John. The Story Gestalt: A model of knowledge-intensive processes in text comprehension. Cognitive Science, volume 16, pages 271-306, 1992.

[St. John 1992b] Mark F. St. John and J.L Mc-Clelland. Parallel constraint satisfaction as a comprehension mechanism. In: R.G. Reilly and N.E. Sharkey (Eds.), Connectionist approaches to natural language processing, pages 97-136, 1992.

[Svore 2007] Krysta Svore, Lucy Vanderwende, Christopher Burges. Enhancing single-document summarization by combining RankNet and third-party sources. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 448-457, 2007.

[TAC 2008] Proceedings of the Text Analysis Conference (TAC). NIST, pp. 17-19. Gaithesburg, Maryland, USA, 2008.

[Tomer 2021] Minakshi Tomer, Manoj Kumar, Multi-document extractive text summarization based on firefly algorithm, Journal of King Saud University - Computer and Information Sciences, 2021.

[Torres-Moreno 2015] Juan-Manuel Torres-Moreno. Trivergence of probability distributions, at glance. arXiv, 2015.

[Torres-Moreno 2014] Juan-Manuel Torres-Moreno. Automatic Text Summarization.

Wiley publishing, 2014.

[Torres-Moreno 2010] Juan-Manuel Torres-Moreno, Horacio Saggion, Iria da Cunha, Eric SanJuan, and Patricia Velázquez-Morales. Summary evaluation with and without references. Polibits, (42) :13–20, 2010.

[Trieu 2020] Nicholas Trieu, Sebastian Goodman, Pradyumna Narayana, Kazoo Sone and Radu Soricut. Multi-Image Summarization: Textual Summary from a Set of CohesiveImages. abs/2006.08686. 2020.

[Trabasso 1985] Tom Trabasso and Linda L. Sperry. Causal relatedness and importance of story events. Journal of Memory and Language, volume 24, pages 595-611, 1985.

[Trandabâț 2011] Diana Trandabâµ. Using semantic roles to improve summaries. In Proceedings of the 13th European Workshop on Natural Language Generation, pages 164–169, Nancy, France, September 2011.

[Usunier 2005] Nicolas Usunier, Massih-Reza Amini and Patrick Gallinari. Résumé automatique de texte avec un algorithme d'ordonnancement. In Proceedings of the 2nd Conférence en Recherche d'Informations et Applications (CORIA 2005), pages 25–40, 2005.

[Van Den Broek 1996] Paul van den Broek, Kirsten Risden, Charles R. Fletcher and Richard. Thurlow A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In: B.K. Britton and A.C. Graesser (Eds.), Models of Understanding Text, Mahwah, NJ: Erlbaum. pages 165-187, 1996.

[Van-Dijk 1983] Teun A. van Dijk and Walter Kintsch. Strategies of Discourse Comprehension. Academic Press. New York, 1983.

[Van-Halteren 2003] Hans Van-Halteren and Simone Teufel. Examining the consensus between human summaries: initial experiments with factoid analysis. In Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5, pages 57–64. Association for Computational Linguistics, 2003.

[Vishal 2021] Gupta Vishal, Dilip K. Sharma, and Ashutosh Dixit. Review of Information Retrieval Models, Performance Evaluation Techniques and Applications. International Journal of Sensors, Wireless Communications and Control 11. 2021.

[Wei 2018] Huawei Wei, Bingbing Ni, Yichao Yan, Huanyu Yu, Xiaokang Yang and Chen Yao. Video Summarization via Semantic Attended Networks. In AAAI. 2018.

[Widyassari 2019] Adhika P. Widyassari, Affandy Affandy, Edy Noersasongko, Ahmad Zainul Fanani, and Ruri Suko Basuki. Literature Review of Automatic Text Summarization: Research Trend, Dataset and Method. In International Conference on Information and Communications Technology (ICOIACT), pqges 491-496, Yogyakarta, Indonesia, 2019.

[Widyassari 2020] Adhika P. Widyassari, Supriadi Rustad, Guruh F. Shidik, Edi Noersasongko, Abdul Syukur, Affandy Affandy, and Ignatius M. Setiadi. 2020. Review of Automatic Text Summarization Techniques & Methods. Journal of King Saud University - Computer and Information Sciences. 2020.

[Woodsend 2010] Kristian Woodsend, Mirella Lapata. Automatic Generation of Story Highlights, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 565–574, 2010.

[Zhan 2009] Jiaming Zhan, Han Tong Loh et Ying Liu. Gather customer concerns from online product reviews - A text summarization approach. Expert Syst. Appl., vol. 36, no. 2, pages 2107–2115, March 2009.

[Zhang 2019] Junsheng Zhang, Kun Li, Changqing Yao and Yunchuan Sun. Event-based summarization method for scientific literature. J. Personal and Ubiquitous Computing. Volume 1, issue 10. 2019.