

- 36- إعجاز القرآن بين المعتزلة والأشاعرة، منير سلطان، ص56.
- 37- فكرة إعجاز القرآن، نعيم الحمصي، ص36.
- 38- أنظر إعجاز القرآن بين المعتزلة والأشاعرة، منير سلطان، ص52/51 بتصرف.
- 39- إعجاز القرآن، الباقلاني، ص08.
- 40- شرح الأصول الخمسة، القاضي عبد الجبار، ص152.
- 41- إعجاز القرآن، الباقلاني، ص08.
- 42- المعجم العربي نشأته وتطوره، حسن نصار، ج1 ص27/26. دار النصر للطباعة 1988م.
- 43- أثر القرآن في تطور النقد العربي، ص38/37.

LES RESUMES ET LES RESUMES AUTOMATIQUES

Aidjouli Hocine

Centre universitaire de Tissemsilt

Cet article abordera le domaine des résumés par un ensemble de définitions et de concepts liés à la réduction des textes que ce soit manuel ou automatique.

A/ Les résumés

1. Les définitions :

Dans la littérature, Il existe plusieurs définitions du terme «résumé», parmi ces différents définitions on trouve celle du colloque international du résumé des sciences qui s'est tenu à Londres entre le 20 et le 25 du mois de juin 1949 et qui stipule que :

« Le résumé est une contraction d'un ouvrage ou de documents accompagnée d'une description sur un support comportant la facilité d'accès au document d'origine »⁽¹⁾.

Alain conte⁽²⁾ stipule que le résumé est une contraction d'un texte accompagné d'une bibliographie permettant l'accès à ce texte.

Mohamed Fathi Abdelhadi⁽³⁾ définit le résumé comme étant une présentation restreinte et exacte du contenu d'un document dont le style ressemble à celui-ci, accompagné d'une description bibliographique permettant l'accès au premier document.

Horacio Saggion de l'université de Montréal⁽⁴⁾, définit le résumé comme étant un texte concis qui rend compte du contenu « essentiel » d'un document.

Plus généralement, Gergisse⁽⁵⁾ et Abdeljabar Abderrahmane définissent le résumé comme une sorte de forme bibliographique qui donne le même sens qu'un texte ou un document, et qui permet d'émettre la même idée sur le secteur de traitement de ces documents facilitant ainsi la recherche suivie.

Résumer c'est donc transformer un texte en un texte plus court mais de sens équivalent (fidélité aux idées du texte original)

2. Apparition des résumés:

Les résumés sont apparus selon Bradford⁽⁶⁾ au commencement d'édition des premiers périodiques relatifs à ceux-ci à partir de Janvier 1665. Toutefois les résumés sont apparus bien avant⁽⁷⁾, mais l'apparition réelle revient à l'édition des résumés chimiques de l'association américaine de la chimie au 19^{ème} siècle.

3. L'importance des résumés :

L'importance des résumés comme outil d'accès à l'information réside dans ce qui suit :

- Les résumés sont considérés comme un style évolué permettant aux chercheurs le suivi des évolutions des sciences sans rentrer dans les détails, vu l'explosion du secteur de l'information.
- Le bon résumé aide à l'économie de l'effort et du temps du lecteur du fait qu'il fait gagner environ 9/10 du temps qui serait normalement dégagé pour la lecture du document d'origine.
- Le résumé aide à l'économie des frais de recherches sur un sujet.
- Le résumé aide à éviter les problèmes de traduction du fait de l'existence d'environ 70 langues au niveau des sciences et de la technologie. Un chercheur qui maîtrise en moyenne deux langues peut éviter ce genre de problème en utilisant la langue la plus répandue au niveau des résumés.
- Les résumés aident à faire le choix des documents à utiliser relativement à un certains sujets, du fait que certains titres de documents peuvent être très généralistes ceci peut être dû à la sécurité par exemple ou le mauvais choix de titre.
- Les résumés aident aussi à la recherche d'informations pertinentes sur un sujet.
- Les résumés aident au développement des sciences du fait de leur pertinence dans le style et de l'économie de l'effort, du temps et des frais de recherches.
- Les résumés aident à faciliter la révision et le contrôle de la documentation scientifique du fait de leur simplicité par rapport aux documents d'origines.

Additionnement à cela, Cleber Guenitcha et Michel Manou⁽⁸⁾ stipulent que les résumés ont trois fonctions importantes :

1. La diffusion des informations
2. L'acquisition de l'information pertinente
3. Restitution de l'information

De son côté Rafik El-halimi⁽⁹⁾ stipule l'importance des résumés dans ce qui suit :

1. nous ramène à développer la lecture critique et la compréhension totale de ce qu'on lit.
2. Nous aide à éviter la perte du temps et de l'effort en lisant des détails non trop important.
3. On distingue les idées principales et secondaires.
4. On ingurgite rapidement un nombre important d'idées dans un monde d'explosion de l'information.
5. Nous permet de voir le sens esthétique des textes et leurs différents types.
6. Nous aide à être sûrs de nous-mêmes car c'est un effort individuel.
7. Nous est utile lors des études, des conférences ou dans la recherche scientifique.

4. Les méthodes des résumés :

Les règles les plus connues et les plus utilisées en résumé sont les suivantes :

1^{ère} règle : La suppression

On peut supprimer totalement les phrases qui ne contribuent pas à la compréhension du texte. A titre d'exemple : la description des objets, d'individus ou de tâches secondaires.

2^{ème} règle : La fusion

On peut fusionner une phrase au niveau d'autres qui sont des conditions essentielles de celle-ci ou des conséquences.

3^{ème} règle : La construction

On peut construire une phrase à la place d'autres à condition qu'elle soit le résultat naturel de celles-ci.

4^{ème} règle : La généralisation

On peut substituer des phrases par une autre plus globale ayant le même sens.

Le fait de résumer n'est pas la seule caractéristique des matières scientifiques, littéraires ou écrites, c'est aussi nécessaire dans l'ère de l'information et de l'explosion des connaissances.

En effet, on peut en avoir besoin pour reformuler des matières audio-visuelles en effaçant ce qui est superflus et en ne gardant que l'essentiel par un traitement aboutissant à des idées simples et claires⁽¹⁰⁾.

5. Les types de résumés :

Il existe plus d'un critère pour la typologie des résumés, en effet, il est possible de classer ces dernières selon le but de leur élaboration en: résumés informatifs, indicatifs, descriptifs ou critiques; ou encore selon la méthodologie suivie pour résumer en résumés unifiés, télégraphiques ou adaptés.

Alain Conte classe les résumés comme suit

1. Les résumés traditionnels qui englobent les résumés descriptifs et les résumés informatifs

2. les résumés adaptés

3. les résumés de style unifié

Alors que Mohamed Mohamed Amani⁽¹¹⁾ les classe de la façon suivante :

1. Le résumé informatif

2. Le résumé descriptif

3. Le résumé mixte

4. Le résumé prévisionnel

5. Le résumé d'auteur

Atefeh Farzindar de l'université de Montréal⁽¹²⁾, cite les types de résumé suivants :

Résumé indicatif: fournit une idée du texte sans donner le contenu spécifique, il signale les thèmes du document.

Résumé informatif : renseigne sur les informations essentielles contenues dans le document.

Résumé sélectif: qui néglige les aspects très généraux d'un document et développe les parties spécialisées.

Résumé cible: se concentre sur le(s) topique(s) d'intérêt pour l'utilisateur.

Résumé générique: reflète le point de vue de l'auteur.

6. Les documents concernés par le résumé :

D'après Owlley Jennifer⁽¹³⁾, les documents concernés par le résumé sont les suivants :

1. Les documents ciblés par les utilisateurs

2. Les documents qui apportent du nouveau dans les différents domaines critiques

3. Les rapports finaux qui sont argumentés

4. Les documents qui contiennent des informations difficiles à obtenir ou des informations rares

5. Les présentations importantes

6. Les magazines de renommée

En outre, Hechmat Kacem ajoute les classes suivantes:

1. Les manuels d'utilisation ainsi que les brevets

2. Les recherches scientifiques aboutissant à de nouveaux résultats

3. Les thèses de recherche

B/ Les résumés automatiques

Le traitement automatique des langues (TAL) est une branche de l'intelligence artificielle, son objectif est la conception de logiciels ou programmes, capables de traiter de façon automatique des données linguistiques, c'est-à-dire des données exprimées dans une langue dite 'naturelle'⁽¹⁴⁾.

Parmi les diverses applications possibles qui peuvent être rattachées au TAL, on peut citer principalement⁽¹⁵⁾:

- **La traduction automatique** : impliquant compréhension d'une langue, transfert dans une autre langue et génération dans cette dernière.
- **Le résumé automatique des textes** : l'analyse est suivie d'une étape de sélection des éléments considérés comme essentiels et leur génération.
- **La réalisation d'interface en langue naturelle** : permet l'interrogation de bases de données, l'utilisation de systèmes experts ou la constitution de leur base de connaissances. On a alors surtout besoin de connaissances sur le domaine (souvent assez restreint, alors que les deux types d'applications précédents, le domaine est ouvert) plus que de connaissances générales et de procédures de gestion de dialogue élaborés.
- **L'indexation automatique de documents, et la recherche documentaire** : Plus que de compréhension, il s'agit ici de projection des documents traités sur un certain nombre de concepts clefs plus ou moins fixés à priori, en vue de retrouver les plus pertinents à partir d'une question donnée.
- **La commande de robot** en langue naturelle demandera bien entendu une compréhension bien détaillée des ordres donnés, ainsi qu'une bonne représentation du monde dans lequel se trouve le robot. Des problèmes de planification devront être abordés, et selon le type de tâche envisagée, des procédures de dialogue pourront aussi être nécessaires.
- **L'enseignement assisté par ordinateur** : Là, une représentation fine de l'utilisateur est nécessaire, afin que le programme puisse s'adapter au profil particulier de l'élève qu'il a devant lui. Des procédures de gestion de dialogue et un module pédagogique élaborés seront alors essentiels, alors que les connaissances sur le monde seront souvent assez réduites.

1. Le résumé automatique

Un résumé est un texte concis qui rend compte du contenu «essentiel» d'un document. Comment faire pour qu'un ordinateur puisse calculer le contenu essentiel d'un document et l'exprimer sous la forme d'un nouveau texte cohésif et cohérent? C'est la problématique du résumé automatique.

La notion de **résumé automatique** devient un des grands thèmes du Traitement Automatique des Langues. Plutôt que de diffuser les documents entiers, n'est-il pas préférable de diffuser seulement les résumés qui contiendraient les informations vraiment pertinentes? En effet, il est plus facile de lire quelques lignes ou quelques pages susceptibles d'apporter l'information cherchée que de lire des centaines de pages pour s'apercevoir qu'aucune information nouvelle ne s'y trouve.

2. Les approches de résumé automatique:

Il existe deux approches principales pour produire un résumé automatique:

1. Produire un **extrait** du document source

2. Ou l'approche **hybride** formée de plusieurs méthodes d'extraction .

2.1 Les méthodes d'extraction de phrases

L'objectif des méthodes d'extraction des phrases est de repérer dans le texte source les phrases les plus importantes. Le résultat obtenu est alors un extrait du texte source⁽¹⁶⁾.

2.1.1 Méthodes à base de mots clés

Cette méthode est basée sur le fait que l'auteur se sert (pour exprimer ses idées principales) de quelques mots-clés qui ont tendance à être récurrents dans le texte⁽¹⁷⁾. Le résumé automatique est alors produit en recherchant dans le texte source les unités de texte minimales réunissant ses mots-clés. Ce principe est souvent appliqué en différentes variantes présentées dans les sous-sections qui suivent.

2.1.1.1 Mots-clés prédéfinis

Pour calculer le score de chaque phrase S selon les mots-clés qu'elle contient, on peut calculer le score suivant :

$$\text{Score}_{\text{mot-cle}}(S) = \sum_{w \in S} a(w) \times F(w)$$

$$\text{Où } a(w) = \begin{cases} A & \text{si } w \in \text{liste de mot-clés } (A > 1) \\ 1 & \text{sinon} \end{cases}$$

et $F(w)$ est la fréquence du terme w dans la phrase S

La liste de mots-clés peut être introduite par l'utilisateur (domaine d'intérêt) ou composée des mots-clés établis par l'auteur. L'importance du poids du terme w est donné par $A \times F(w)$, avec $A > 1$.

2.1.1.2 Titres

Étant donné que le titre est l'expression la plus significative et qui résume le mieux un document en quelques mots, on peut dire que la phrase qui ressemble le plus au titre est la plus marquante du document. Par conséquent, on peut attribuer à chaque phrase un poids en fonction de sa ressemblance avec le titre⁽¹⁸⁾.

Dans ce cas on considère les mots du titre du texte comme des mots-clés et on produit le résumé en sélectionnant les phrases qui couvrent certains mots apparaissant dans un titre.

2.1.1.3 Méthode de distribution de termes (DT)

Le pionnier dans le domaine de la génération de résumés est Luhn⁽¹⁹⁾ qui a introduit la méthode de distribution de termes. L'idée de cette méthode est de considérer comme «importantes» les phrases qui contiennent des mots «importants» du texte. Un mot est considéré important s'il est employé assez fréquemment dans le texte.

En premier lieu, le texte source est traité pour calculer la fréquence de chaque mot de «contenu» du texte et en second lieu les fréquences sont utilisées pour associer un poids à chaque phrase.

On considère comme un même mot les mots dérivés de la même racine (par exemple, résumé, résumés et résumer).

Une fois la fréquence de chaque mot calculée, une liste triée par fréquence est obtenue, il s'agit de la liste de distribution de termes.

Pour mesurer le poids d'une phrase, on utilise le texte source et la liste de distribution de termes.

Ensuite, on choisit les phrases les plus « pesantes ». La sélection peut être faite en terme d'un pourcentage du texte original, en nombre de phrases ou en nombre de mots.

La génération du résumé consiste à juxtaposer les unités sélectionnées en ordre d'apparition dans le texte source.

Les avantages de la méthode sont sa robustesse (n'importe quel texte aura un résumé) et sa facilité d'implantation. Les limitations sont toutefois nombreuses. Comme on ne prend pas en considération les relations entre les différents éléments du texte, le résultat risque d'être incohérent et même d'omettre de l'information importante.

2.1.2 Méthode de la position (P)

Cette méthode a été introduite par Edmunson⁽²⁰⁾ pour compléter la méthode de distribution de termes. Elle est utilisée en combinaison avec d'autres méthodes d'attribution de poids pour faire augmenter ou diminuer le poids d'une phrase.

La méthode de la position considère que les premières et dernières phrases de chaque paragraphe sont importantes car elles sont considérées comme thématiques, c'est-à-dire elles «résumement» le contenu du paragraphe, donc ces phrases auront leurs poids augmentés (cette affirmation est appuyée par les expériences de Baxendale⁽²¹⁾). La méthode considère aussi des phrases positionnées dans certaines sections conceptuelles importantes, par exemple dans «Introduction» et «Conclusion».

L'inconvénient de cette méthode est qu'elle dépend de la nature du texte à résumer ainsi que du style de l'auteur.

2.1.3 Expressions indicatives

Cette méthode a été introduite par Paice⁽²²⁾ avec l'objectif de produire des résumés indicatifs. Dans la littérature scientifique, on trouve souvent des expressions qui, indépendamment du domaine particulier du texte, font référence à des catégories conceptuelles. Ainsi, si dans l'introduction d'un article on trouve une phrase avec l'expression «L'objectif de cet article est...», on peut être presque sûr que ce qui suit dans la phrase est l'information sur les objectifs de l'article et lorsqu'on trouve une phrase qui commence par «Pour en conclure...», on a la certitude qu'il s'agit d'une conclusion.

L'idée de la méthode des expressions indicatives est de sélectionner dans tout le texte des phrases contenant ces types d'expression. Chaque expression indicative pourra avoir un poids associé et, alors, la sélection des phrases sera basée sur leur poids tel que fait Lehman⁽²³⁾.

2.1.4 Classification des éléments

Dans les textes de science et de technique il y a des phrases qui font référence à des catégories conceptuelles telles que: Connaissances Antérieures, Contenu, Méthode et Résultat, on peut également constater que dans les résumés de science et technique des informations relatives à ces catégories sont souvent retenues pour le résumé.

Cette approche essaie de classifier sémantiquement les phrases d'un texte tout en oubliant le contexte⁽²⁴⁾.

2.2. Approches hybrides

Les méthodes présentées dans les sections précédentes utilisent des traits (fréquence, position, expression indicative, etc.) qui ne peuvent isolément garantir des résultats optimaux. On combine souvent ces traits par exemple avec l'équation suivante :

$$\text{Score}_{\text{hybride}}(S) = a_1 * \text{Score}_{\text{DT}}(S) + a_2 * \text{Score}_{\text{P}}(S) + a_3 * \text{Score}_{\text{exp.ind}}(S) + a_4 * \text{Score}_{\text{Titre}}(S)$$

Les poids a_i peuvent être fixés arbitrairement ou déterminés de manière expérimentale (par apprentissage par exemple).

Dans le cas de textes journalistiques, Strzalkowski et al.⁽²⁵⁾, ont combiné les méthodes de distribution de termes, du titre et de la position, en considérant la spécificité du texte. Pour garder la cohérence du texte, le résumé est composé d'une sélection de paragraphes pertinents.

3. Les domaines d'applications des résumés automatiques:

Nombreux sont les domaines d'outils du résumé automatique, en effet, on peut recenser les usages suivants:

- Pour résumer les nouvelles au SMS pour les téléphones portables.
- Pour laisser un ordinateur synthétique lu le texte résumé. Le texte écrit peut être long et ennuyeux pour le lire.
- Dans des moteurs de recherche pour présenter les descriptions compressées des résultats de recherche.
- Pour chercher dans des langues étrangères et obtiennent un résumé automatiquement traduit du texte automatiquement résumé.

Sans oublié les domaines de l'archivage, des bibliothèques et du journalisme.

CONCLUSION

Dans cet article, nous avons passer en revue, de ce qui a trait au domaine du résumé automatique, ses définitions et ses techniques.

Nous avons répertorié toute les méthodes d'extraction, hybrides.

Les méthodes d'extraction offrent beaucoup d'avantages. Certaines méthodes semblent offrir de meilleurs résultats que d'autres, cela est dû en grande partie à la nature du texte et au style de l'auteur.

L'approche mixte (hybride) s'apprête mieux utilisée: vu que les coefficients des paramètres peuvent être déterminés expérimentalement, on arrive à obtenir de bons résultats.

Bibliographie

- 1- (1999) غالب عوض النوايسة، الاستخلاص لأغراض استرجاع المعلومات، مجلة رسالة المكتبة، الأردن
- 2- Alain Conte, La révolution d'informations : l'utilisation des ordinateurs pour le stockage et la récupération d'informations, traduction Hechmat Kacem et Chawki Salem 2^{ème} édition, Kuwait : l'agence des publications, p172, 1979.
- 3- محمد فتحي عبد الهادي، العمليات الفنية في المكتبات ومراكز البحث (1983) المجلة العربية للمعلومات، مج4، ع2، ص44
- 4- Horacio Saggion, Génération automatique de résumé par analyse sélective Thèse de doctorat, Université de Montréal 2000.
- 5- جرجيس جاسم محمد، عبد الجبار عبد الرحمان، المراجع والخدمات المرجعية مركز التوثيق ص32، (1985) لدول الخليج العربي الإعلامي.
- 6- قاسم حشمت، خدمات المعلومات: مقوماتها وأشكالها، القاهرة: مكتبة غريب، ص209 (1984)

- 7- كتاب التلخيص في وجوه البلاغة للإمام جلال الدين محمد بن عبد الرحمن القزويني ضبط وشرح عبد الرحمن البرقوقي دار الفكر العربي 1999.
- 8- Cleber Guenitcha et Michel Manou, Introduction générale des sciences et techniques d'informations et d'archivage, UNISCO, p171, Tunis, 1987.
- 9- رفيق حسن الحليمي، فن كتابة التلخيص والمختصرات، مركز المخطوطات والتراث والوثائق 2002 الكويت ط 1.
- 10- 2006 شفيق سالم الشاعر. معلم بوكالة الغوث. فن التلخيص، مجلة المعلم تربوية ثقافية جامعة
- 11- 1985 محمد محمد أمان، خدمات المعلومات والإحاطة الجارية، الرياض، دار المريخ، ص 67
- 12- Atefeh Farzindar, Développement d'un système de résumé automatique de textes juridiques, Université de Montréal 2001.
- 13- Owlley Jennifer, Abstracting and indexing, London : Clive Bingley, p12-20, 1982
- 14- Mourad Amine, Glossaire de linguistique computationnelle 1995.
- 15- Gerard sabah, Le TALN & les modèles de représentation des connaissances 1988
- 16- M. Monod, Le résumé automatique, un petit état de l'art 2004
- 17- T.A.S. Pardo, L.H.M. Rino, M.G.V. Nunes, Extractive summarization: how to identify the gist of a text, *International Bibliographie 63 Information Technology Symposium - I2TS 2002, Florianópolis-SC, Brazil, pp.245-260, 01-05 October 2002.*
- 18- Ishikawa, K., Ando, S., Okumura, A.: Hybrid Text Summarization Method based on the TF Method and the Lead Method. *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization. Tokyo.Japan. pp.5-219-5-224, March 2001.*
- 19- P. H. Luhn, The Automatic Creation of Literature Abstracts, IBM Journal April 1958, pp. 159-165, April 1958.
- 20- H. P. Edmundson: New methods in automatic abstracting, *Journal of the Association for Computing Machinery (ACM), vol. 16 N°2 pp. 264-285, April 1969.*
- 21- J. Baxendale, Man-made Index for Technical Literature - an experiment. IBM J. Res. Dev., 2(4) :354-361, 1985.
- 22- D. Paice, The Automatic Generation of Literary Abstracts : An Approach based on Identification of Self-indicating Phrases. In Norman, O., Robertson, S., van Rijsbergen, C., and Williams, P., editors, Information Retrieval Research, Butterworth, London 1981.
- 23- A. Lehman, Automatic Summarization on the WEB. A System for Summarizing using Indicating Fragments : RAFI. In Proceedings of Computer-Assisted Information Searching on Internet Conference. RIAO'97, pages 112-122, McGill University, Quebec, Canada, 1997.
- 24- Larkey, L. S., Ballesteros, L. and Connell M., improving Stemming for Arabia Information Retrieval: Light Stemming and co-occurrence Analysis, In proceeding of the 25th annual International conference on Research and development in information Retrieval (SIGIR 2002), Tampere, Finland, August 2002.
- 25- T. Strzalkowski, J. Wang and B. Wise, Summarization-based Query Expansion in Information Retrieval, *Proceedings of 36th Annual Meeting of the ACL, V. 2, pp. 1258-1264, Montreal 1998.*