UNIVERSITE DE DROIT ET DES SCIENCES D'AIX-MARSEILLE
Faculté des Sciences et Techniques de Saint Jérôme

LA SERIATION DES SIMILARITES SPECIFIQUES : OUTIL POUR LA RECHERCHE DE L'INFORMATION STRATEGIQUE

Une méthode de classification automatique de l'information issue des bases de données en Veille Technologique

THESE
Présentée et soutenue publiquement par :

BALDIT Patrick

Ingénieur ESCM
Inghieur ESIPSOI

le

pour obtenir le grade de
Docteur Ingénieur

Spécialité :
Sciences de l'information et de la communication

Président de Jury :	Marcotorchino F. (Professeur Paris VI)
Autres membres du jury :	Paoli C. (Professeur Marne La Vallée)
	Ruiz J.M. (Professeur ENSSPICAM)
	Dou H. (Professeur CRRM)
	Cornu L. (Professeur Université Aix-Arseille II)
	Quoniam L. (Maître de Conférence CRRM)

Rapport concernant le mémoire de doctorat présenté par Patrick Baldit.

La sériation des similarités spécifiques; Outil pour la recherche de l'information stratégique.

Le mémoire présenté comprend trois parties.

Le premier chapitre est consacré à l'analyse du contexte de la réalisation du candidat. Cette analyse du contexte Veille Technologique et Intelligence Economique est elle même replacée dans l'environnement le plus global de la gestion de projets. Le vocabulaire est ici bien défini ainsi que les concepts, et **les** méthodes. Cette partie replace notamment les activités de Bibliométrie pour le traitement de corpus volumineux afin de valoriser les références bibliographiques dans le cadre d'activités de Veille Stratégiques. Le candidat montre comment l'analyse des informations accessibles sur les réseaux à hauts débits peut aider à la recherche d'innovations indispensables aux activités de Veille Technologiques et Intelligence Economique. Cette partie fait un point sur la production d'information stratégique pour l'aide à la décision.

Le deuxième chapitre rappelle l'environnement technique de la réalisation **du** candidat. Les lois bibliométriques sont rappelées ainsi que les différentes méthodes d'analyse de données. II introduit la nécessité d'outils mieux adaptés aux données à traiter. Cette adaptation va dans le sens de plus gros volumes traités et de respects de propriétés spécifiques. Ce chapitre montre également l'intégration avec les outils déjà existant dans au CRRM où le candidat a fait sa thèse.

La troisième partie constitue l'essentiel du mémoire. Elle présente le produit DATABLOC développé au CRRM dans le cadre de ce doctorat. La présentation va du cadre théorique de la méthode de sériation par blocs aux applications en passant par une présentation du produit réalisé. Les capacités de traitement de ce nouveau produit et son interface conviviale ouvre des perspectives importantes en matière de traitement de grandes bases de données bibliographiques. L'application présentée, au niveau de l'exploitation analytique d'enquêtes par interview, montre que le candidat sait interconnecter des domaines apparemment éloignés et comment cet outil peut servir d'aide à la lecture de textes.

L'exposé, de la démarche, est très clairement exprimé et le manuscrit est bien structuré.

Cette thèse a le grand mérite de replacer des outils mathématiques dans un contexte global, mais surtout de les valider par la réalisation d'un logiciel opérationnel et par son application à des cas concrets.

L'ensemble des travaux effectués par Patrick Baldit mérite sans aucun doute d'être présenté en vue de l'obtention d'un doctorat.

Fait à Marseille le 26110194

Lucienne CORNU
Professeur à l'université Aix-Marseille II

A Maïté mon épouse
ef Benjamin mon fils

Je tiens à remercier en tout premier lieu mon laboratoire d'accueil, le Centre de Recherche Rétrospective de Marseille, et plus particulièrement Monsieur **Henri Dou** qui a su mettre à ma disposition un matériel performant et des conditions de travail propice au déroulement optimal de mes travaux.

Je tiens également à saluer ici **Christophe RETOURNA et Raimundo DOS SANTOS,** qui m'ont donné à maintes reprises l'occasion d'étayer mon travail notamment dans la phase de mise au point du logiciel DATABLOC.

Je remercie l'ensemble du laboratoire d'ingéniérie des projets industriels en la personne de **Jean Michel RUIZ,** qui m'a permis d'entreprendre des développements concernant des applications hypertextes en enseignement assisté par ordinateur

Je remercie également les **membres du jury** et les **rapporteurs** de ma thèse de bien avoir voulu consacrer leur temps à la lecture de ce mémoire

Je ne pourrais terminer sans remercier **Luc Quoniam,** sans lequel ce travail ne serait pas ce qu'il est, et en qui j'ai trouvé une disponibilité de **tous les instants.** J'ai pu apprécier en lui une compétence et un soutien moral qui m'ont été précieux. Qu'il en soit ici chaleureusement remercié.

Je garderais enfin une pensée pour mon épouse Maïté et mon fils Benjamin qui m'ont permis par leur collaboration de pouvoir achever cette thèse dans de bonnes conditions.

Résumé de la thèse

Ces dernières années ont vu l'émergence du concept d'intelligence économique sensibilisant les décideurs publics ou privés à se préoccuper des modifications de **leur** environnement. Depuis une trentaine d'années, au niveau de la planète, le stockage informatique de données couvrant un vaste domaine de connaissance s'est généralisé. La problématique s'est orientée vers une exploitation rationnelle de ce gisement mondial dans le but d'en retirer des informations pertinentes et stratégiques. L'utilisation de la bibliométrie comme outil de traitement dans le cadre de la veille technologique s'est imposée pour permettre d'établir des grilles de lecture de documents primaires pour des experts du domaine. Les distributions spécifiques de ce type d'information rendent l'utilisation des techniques statistiques usuelles difficilement exploitables et nous ont conduit à la création d'un algorithme de traitement adapté, permettant une classification non hiérarchique par optimisation d'un critère global, que nous avons appelé la sériation des similarités spécifiques. L'objectif final du traitement statistique étant de fournir aux experts des représentations cartographiques de l'information recueillie, le développement de visualisation hypertexte s'est avéré primordial pour une exploitable conviviale.

Mots Clefs

Information Stratégique, Analyse de Donnée, Veille Technologique, Base de Données, Indices d'association, Brevets, Classification Automatique, Bibliométrie, Bloc Sériation, Classification Non Hiérarchique.

Sommaire

Sommaire

Liste des figures

Chapitre II : L'information stratégique

Chapitre III : La bibliométrie et l'analyse de données