
Évaluation de la pertinence des résultats en recherche d'information géographique

Définition d'un cadre expérimental et validation de l'apport des dimensions de l'information géographique

Guillaume Cabanac* — **Damien Palacio****
Christian Sallaberry** — **Gilles Hubert***

* Université de Toulouse, IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9

** Université de Pau et des Pays de l'Adour, LIUPPA ÉA 3000
Avenue de l'Université, BP 1155, F-64013 Pau cedex

prenom.nom@irit.fr; prenom.nom@univ-pau.fr

RÉSUMÉ. L'étude des requêtes soumises aux moteurs de recherche montre une utilisation croissante des dimensions spatiale et temporelle en complément de la thématique seule. Des travaux proposent d'intégrer ces trois dimensions de l'information géographique dans le processus de recherche d'information (RI). Leur qualité n'est actuellement mesurée qu'au travers d'indicateurs liés au stockage et au temps de réponse. Cependant, la pertinence des résultats n'est pas quantifiée. Dans ce contexte, notre proposition est triple. Premièrement, nous définissons un cadre d'évaluation de la qualité de tels systèmes. Deuxièmement, nous l'exploitons pour tester l'hypothèse suivante : la combinaison des trois dimensions améliore la qualité des résultats. Troisièmement, nous montrons la complémentarité de ces dimensions pour le processus de RI.

ABSTRACT. The analysis of search engine logs reveals an increasing use of spatial and temporal dimensions besides the topical dimension. The integration of these three geographic dimensions into the information retrieval (IR) process has been addressed. Measurement of their quality focuses on efficiency (index sizes and query times). Their effectiveness (i.e., accuracy of results) is not quantified, however. Our proposal is threefold. First, we define a framework for measuring the effectiveness of Geographic IR systems. Second, we experiment it for testing this assumption: combining the three dimensions leads to an improvement. Third, we show that these complement each other, making them worth using in the IR process.

MOTS-CLÉS : Recherche d'information géographique, évaluation, combinaison de résultats.

KEYWORDS: Geographic Information Retrieval, evaluation, retrieval combination.

1. Introduction

La proportion de requêtes géographiques soumises aux moteurs de recherche généralistes a fait l'objet d'études récentes. Ainsi, pour Excite (Sanderson *et al.*, 2004), AOL (Gan *et al.*, 2008) et Yahoo (Jones *et al.*, 2008), cette proportion varie entre 12,7 % et 18,6 %. Or, bien que les moteurs généralistes restituent déjà de bons résultats pour des recherches par mots-clés, Kanhabua *et al.* (2008) ont observé que, sur des corpus spécifiques, la précision des recherches géographiques est faible. De fait, l'utilisateur passe beaucoup de temps à explorer les documents restitués afin de ne retenir que ceux qui satisfont réellement son besoin. Par exemple, au sein d'une requête plus large, l'expression temporelle « *les années 1810* » soumise à un moteur de recherche classique conduit à la restitution de documents contenant « *1810* » et non « *1811* », « *1812* », etc. De même, l'expression spatiale « *autour d'Anglet* » soumise à un moteur de recherche classique conduit à la restitution de résultats contenant « *Anglet* » et non « *le golf de Chiberta* », « *la plage des Cavaliers* », « *Bayonne* », etc. Un moyen d'améliorer l'efficacité des moteurs de recherche est alors d'y inclure la prise en compte des aspects géographiques. Nous reprenons la théorie selon laquelle l'information géographique comporte trois dimensions : spatiale, temporelle et thématique (Gaio, 2001). L'extrait de texte « *Les villes et les châteaux fortifiés dans le bassin aquitain au XIII^e siècle* » illustre bien ce triptyque.

Parallèlement à ces préoccupations observées sur le Web, la numérisation de documents progresse, produisant un volume croissant de documents électroniques. Tandis que certains projets visent simplement la création de versions numérisées de documents, des efforts centrés sur des domaines spécifiques ont des objectifs bien plus ambitieux. Par exemple, afin d'exploiter leurs contenus, les documents sont annotés et indexés conformément à des modèles de données dédiés à la description de domaines particuliers. Ainsi, les fonds documentaires patrimoniaux mobilisent de gros efforts de numérisation. Leur valorisation se fait ensuite via des outils de gestion documentaire classiques (notices descriptives, catalogues thématiques) intégrant généralement des moteurs de recherche plein-texte. Dans ce contexte, le projet PIV¹ « Pyrénées Itinéraires Virtuels » (Gaio *et al.*, 2008) consiste à gérer un fonds documentaire de versions électroniques de documents du XIX^e siècle consacré aux Pyrénées et constitué de journaux, romans et récits de voyages. Il s'agit d'un fonds stable (peu de suppressions et de modifications, des insertions régulières de nouveaux documents) et de petite taille (quelques dizaines de milliers de documents). Ce type de ressource est encore très peu connu et son usage limité aux centres des archives et bibliothèques locales. C'est la raison pour laquelle les collectivités locales souhaitent valoriser ces ressources en les diffusant largement et en offrant des services de recherche adaptés.

Pour intégrer de telles considérations géographiques dans le processus de recherche d'information, le projet PIV repose sur le développement de trois chaînes de traitement dédiées à l'indexation et à la recherche d'information spatiale, temporelle et thématique. Ces travaux se situent à la croisée de plusieurs disciplines : Trai-

1. Le projet PIV est soutenu par la Communauté d'Agglomération Pau Pyrénées.

tement Automatique des Langages Naturels (TALN), Systèmes d'Information Géographiques (SIG), Recherche d'Information (RI) et Recherche d'Information Géographique (RIG). Un prototype, correspondant à chaque chaîne de traitement, extrait et indexe des informations issues de documents textuels et propose un moteur de recherche qui, sur des critères spatiaux (Gaio *et al.*, 2008), temporels (Le Parc-Lacayrelle *et al.*, 2007) ou thématiques (Sallaberry *et al.*, 2007), restitue des paragraphes de documents (récits de voyage, par exemple) associés à un score de pertinence. Or, à ce jour, nous avons évalué chaque prototype monodimensionnel indépendamment (Le Parc-Lacayrelle *et al.*, 2007; Palacio *et al.*, 2010). Afin de considérer l'information géographique dans sa globalité, nous présentons dans cet article des propositions d'intégration des trois dimensions dans un modèle unifié. Afin de valider cette approche multidimensionnelle, il est à présent indispensable de l'expérimenter.

La contribution de cet article vise la conception et la mise en place d'un cadre d'évaluation de systèmes de recherche d'information (SRI) géographique. Le cadre d'évaluation que nous proposons a) est destiné à évaluer les SRI spatiaux, temporels, thématiques et toute combinaison de tels SRI ; b) il met en place une collection de test en langue française adaptée à la RIG et, à moyen terme, c) il pourrait mener à la proposition d'une tâche spécifique dans le cadre d'une campagne d'évaluation de type GeoCLEF (Gey *et al.*, 2006).

L'article est structuré comme suit. La section 2 présente une synthèse de l'état de l'art relatif à l'évaluation de SRI, y compris géographiques. La section 3 est consacrée à la description du cadre d'évaluation de SRI géographique que nous proposons. La section 4 détaille une étude de cas : nous évaluons le SRI spatial, le SRI temporel et le SRI thématique du système PIV ainsi que différentes combinaisons de ces SRI. Notre hypothèse est que la combinaison des trois dimensions donne des résultats plus pertinents que chacune prise indépendamment. Cette hypothèse est vérifiée au travers de l'analyse des résultats d'évaluation des SRI. La section 5 fait état des SRI géographique de la littérature que l'on pourrait également évaluer avec le cadre proposé dans cet article. Enfin, la section 6 conclut l'article et discute nos perspectives de recherche.

2. Adéquation des cadres d'évaluation existants pour la RI géographique

Le domaine de la RI est caractérisé par une longue tradition d'évaluation. Un moyen d'évaluer les SRI repose sur la définition d'une « campagne » qui se déroule généralement de la façon suivante :

1) les organisateurs lancent un appel à participation en décrivant les tâches de RI proposées. Par exemple, une tâche de type *ad hoc* vise la restitution d'une liste de documents pertinents pour une requête ouverte. Une tâche de type *question réponse* vise, quant à elle, la restitution de la réponse à une requête précise. Pour la requête « plages d'Anglet » on obtient alors une liste de documents traitant de ce sujet pour la tâche *ad hoc*, à opposer à la liste des noms des plages concernées pour la tâche *question réponse* ;

2) les concepteurs de SRI intéressés s'inscrivent aux tâches de leur choix. Ils sont alors appelés « participants » ;

3) les organisateurs donnent accès à un *corpus* documentaire et à des besoins en information (*requêtes* commentées) ;

4) les participants traitent le corpus, soumettent les requêtes à leur SRI, puis, transmettent aux organisateurs les résultats obtenus (listes de documents triés par pertinence décroissante, le plus souvent) ;

5) les organisateurs déterminent un ensemble de réponses pertinentes pour chaque requête : les *jugements de pertinence*. Ils confrontent ensuite les résultats des participants à ces jugements à l'aide de mesures. Ces dernières produisent un score qui traduit la performance (critère de qualité des résultats) du SRI pour la requête. L'évaluation globale d'un SRI se fait au travers de l'agrégation des mesures qu'il a obtenu pour l'ensemble des requêtes traitées ;

6) les organisateurs publient les résultats des participants et diffusent (généralement sous conditions) la collection de test (composée du corpus, des requêtes et des jugements de pertinence). Cette collection peut par la suite être réutilisée pour évaluer un SRI en dehors du cadre de la campagne.

TREC (Voorhees *et al.*, 2005) est une campagne de référence en RI permettant d'évaluer des SRI au regard de la dimension thématique. Il existe peu de travaux relatifs à l'évaluation des deux autres dimensions de l'information géographique. La dimension temporelle a fait l'objet du cadre d'évaluation TEMPEVAL (Verhagen *et al.*, 2009). Par ailleurs, Bucher *et al.* (2005) ont proposé de considérer deux dimensions simultanément : spatiale et thématique. Cette proposition se retrouve dans la tâche GeoCLEF (Gey *et al.*, 2006) du cadre CLEF (Peters, 2001). Ce dernier a notamment permis, à partir de requêtes géographiques, l'évaluation de SRI thématiques classiques en RI tels que Lemur (Ogilvie *et al.*, 2001), Lucene (Gospodnetić *et al.*, 2005) et Terrier (Ounis *et al.*, 2005), comme rapporté dans (Perea-Ortega *et al.*, 2008).

En termes de caractéristiques, pour ne citer que l'exemple de GeoCLEF 2008 : la collection de test comprend des articles issus de différents journaux (169 477 articles en anglais, 294 809 en allemand et 210 734 en portugais), 25 besoins en information ciblant les dimensions spatiale et thématique. Onze participants ont été évalués avec les mesures *Average Precision* et *Mean Average Precision* (le lecteur pourra se référer à (Manning *et al.*, 2008, ch. 8) pour une définition de ces mesures).

Les travaux de RIG existants (présentés en section 5) ont tout au plus été évalués du point de vue de la taille des index générés et du temps de réponse des SRI géographique. Ces évaluations quantitatives gagneraient à être mises en perspective avec des évaluations qualitatives. Or, à notre connaissance, il n'existe pas de cadre d'évaluation des trois dimensions de l'information géographique d'un point de vue qualitatif. Il est donc impossible de comparer les moteurs de recherche qui s'efforcent de les traiter simultanément. Pour répondre à ce besoin nous proposons donc, dans la section suivante, un cadre expérimental permettant d'évaluer la RI géographique.

3. Proposition d'un cadre d'évaluation dédié à la RI géographique

Le cadre expérimental proposé s'attache à capitaliser le savoir-faire existant (issu notamment de TREC et GeoCLEF) tout en intégrant les spécificités manquantes relatives à l'information géographique. Aussi, la section 3.1 détaille la constitution d'une collection de test couvrant les trois dimensions, puis la section 3.2 expose l'analyse des résultats de SRI permettant de comparer leur efficacité.

3.1. Constitution d'une collection de test pour évaluer la RI géographique

Dans la littérature, notamment dans le cadre TREC (Harman, 2005), une collection de test comprend trois volets :

1) un ensemble de n « *topics* » formulés par des individus, où *topic* est le terme TREC désignant un besoin d'information. Chaque *topic* est au moins caractérisé par un titre, une description et une narration du besoin. Buckley *et al.* (2000) montrent qu'au moins 25 *topics* sont nécessaires pour réaliser des analyses statistiques pertinentes. Notons cependant que le standard de TREC est à 50 *topics*.

2) le *corpus* regroupant plusieurs documents, certains étant pertinents pour les *topics* proposés. Les corpus TREC comprennent plusieurs centaines de milliers de documents au moins (Voorhees *et al.*, 2005).

3) les « *qrels* », terme TREC désignant les *jugements de pertinence*, associant à chaque *topic* l'ensemble des documents pertinents. Étant donné que le corpus est trop volumineux pour être exhaustivement analysé dans le but d'identifier les *qrels*, TREC recourt à la technique du *pooling*. Ainsi, pour chaque *topic*, un *pool* de documents est constitué à partir des 100 premiers documents restitués par chacun des systèmes participant à la campagne d'évaluation, les doublons sont supprimés (opération d'union ensembliste). L'hypothèse est que le nombre et la diversité des SRI contribuant au *pool* permettront de trouver un maximum de documents pertinents. Enfin, un individu appelé « *assesseur* » examine chaque document du *pool* afin d'identifier s'il répond ou pas au besoin d'information spécifié dans le *topic* considéré. Le document est alors qualifié de pertinent ou de non-pertinent.

De telles collections de test ont été mises en œuvre à plusieurs reprises dans des cadres d'évaluation tels que TREC et GeoCLEF. Notons qu'ils ne prennent pas en compte les trois dimensions de l'information géographique. C'est pourquoi nous proposons d'adapter leur constitution pour évaluer la RI géographique, en fournissant :

1) des *topics* couvrant tout ou partie des trois dimensions. Par exemple, un *topic* pourrait avoir pour titre « *Transhumance dans les Alpes au XIX^e siècle* » et pour narration « *Seront considérés pertinents les documents évoquant la transhumance ou les événements rattachés (quotidien du berger en estive) dans le massif des Alpes entre 1800 et 1899* » ;

2) un *corpus* traitant des trois dimensions : l'aspect thématique classiquement considéré est complété par des éléments spatiaux et temporels ;

3) des *qrels* par dimension pour lesquels l'assesseur évalue l'adéquation entre chacune des trois dimensions considérées (thématique, spatiale et temporelle) et le document. Notons que la seule présence des trois dimensions dans le document ne suffit pas à déduire qu'il est pertinent pour la requête. Considérons par exemple le cas d'un document traitant du thermalisme, puis citant « *Gavarnie* » en tant que lieu de naissance du narrateur. Bien que pertinent spatialement, il ne répond pas à la requête « *thermalisme à Gavarnie* ». C'est en raison de ce type de subtilité que l'assesseur doit également évaluer l'adéquation complète (globale) entre la requête et le document.

Concernant le jugement de chaque document, l'assesseur évalue son adéquation avec chacune des trois dimensions. Cette adéquation est actuellement booléenne pour ne pas surcharger les assesseurs ; ce choix rejoint les observations de Bucher *et al.* (2005) qui soulignent que les jugements graduels par dimension sont inutilement complexes à réaliser. À partir des trois jugements booléens et du jugement global également booléen, la valeur de pertinence $v \in \{0; 1; 2; 3; 4\}$ du document est constituée. Cette valeur traduit d'une part le nombre de dimensions pertinentes et d'autre part la pertinence globale. Notons qu'aucune hypothèse n'est faite sur l'importance relative des dimensions : elles sont considérées équitablement.

4) des *ressources géographiques* nécessaires, d'une part, au géoréférencement des entités spatiales et, d'autre part, à l'interprétation des entités temporelles contenues dans le corpus.

Le protocole expérimental détaillé dans la section suivante vise à mesurer l'efficacité des SRI. Ces derniers sont évalués à partir de leurs *runs* : l'ensemble des documents restitués par topic.

3.2. Protocole d'analyse comparative des SRI géographique

La tâche évaluée est une recherche qualifiée de *ad hoc* dans TREC : le SRI répond à un besoin d'information par une liste de documents ordonnée par pertinence décroissante. L'évaluation vise à mesurer l'efficacité relative des SRI suivants :

- SRI monodimensionnel : thématique (Th), spatial (S) et temporel (Te) ;
- SRI bidimensionnels : Th+S, Th+Te et S+Te permettant de mesurer l'apport de chacune des dimensions dans l'efficacité du SRI ;
- SRI géographique combinant les trois dimensions : Th+S+Te.

Pour un *topic* donné, chaque SRI fournit une liste de couples (d, s) représentant le score s de chaque document d restitué. Classiquement, l'efficacité d'un SRI est évaluée grâce aux mesures² *Average Precision* (*AP*) pour chaque topic et *Mean Average Precision* (*MAP*) globalement. Ces dernières requièrent des *qrels* booléens. Or, dans le protocole expérimental proposé, les *qrels* sont graduels afin de représenter les trois dimensions de l'information géographique. Ces deux mesures ne sont donc pas adap-

2. La définition et les propriétés des mesures *AP*, *MAP* et *NDCG* mentionnées dans cette section sont détaillées dans l'ouvrage introductif à la RI (Manning *et al.*, 2008, ch. 8).

tées. C'est pourquoi nous recourons à la mesure de pertinence graduelle *Normalized Discounted Cumulative Gain (NDCG)* proposée par Järvelin *et al.* (2002) et utilisée notamment dans le cadre de la campagne d'évaluation TREC-9 pour la tâche Web caractérisée par des *qrels* graduels (Voorhees, 2001). Cette mesure implémente deux principes. D'une part, les documents très pertinents ($v \rightarrow 4$ dans notre cas) sont plus intéressants que les documents peu pertinents ($v \rightarrow 1$). D'autre part, un document a d'autant moins d'intérêt pour l'utilisateur que son rang est élevé dans la liste de résultats, car il est d'autant moins probable que l'utilisateur accède à ce document-là.

À l'image du protocole d'expérimentation de TREC, nous proposons deux niveaux de granularité d'évaluation d'un SRI : 1) le niveau topic en calculant *NDCG* et 2) le niveau global en calculant la moyenne arithmétique *moyNDCG* des n valeurs de *NDCG*, fournissant ainsi la mesure globale de performance du SRI.

Au niveau global, les n différences observées $\langle m_i^1 - m_j^1, \dots, m_i^n - m_j^n \rangle$ sont rapportées en pourcent (d'amélioration ou de détérioration), où m_s^t représente la valeur de la mesure m obtenue par le système s pour le topic t . La significativité des tests statistiques calculée pour les différences observées est également rapportée : les p -valeurs de significativité sont calculées avec le test t de Student pairé (la différence est calculée entre les paires de valeurs m_i^t et m_j^t) bilatéral (car $\forall t \in [1; n] m_i^t \not\approx m_j^t$). Bien que nécessitant théoriquement une distribution normale des données, Hull (1993) précise que ce test est en pratique robuste aux violations de cette condition. Par ailleurs, Sanderson *et al.* (2005) montrent que ce test est bien plus fiable que d'autres, tel que le test des rangs signés de Wilcoxon. Concrètement, lorsque $p < \alpha$ avec $\alpha = 0,05$ la différence entre les deux échantillons testés est qualifiée de statistiquement significative (Hull, 1993). Plus la valeur p est petite, plus la différence est significative.

4. Étude de cas : évaluation du prototype de SRI géographique PIV

Afin de valider l'hypothèse selon laquelle combiner les trois dimensions améliore la qualité des résultats, nous exploitons le cadre proposé sur la base du prototype PIV. Pour ce faire, la section 4.1 présente ce prototype. Puis, la section 4.2 détaille la mise en œuvre du cadre d'évaluation dans le but de mesurer la qualité des résultats.

4.1. Le SRI géographique PIV

Cette section décrit des éléments de conception relatifs au prototype de SRI géographique PIV. En particulier, ses trois chaînes d'indexation dédiées sont détaillées en section 4.1.1. Par la suite, la section 4.1.2 est consacrée à la description du processus d'interrogation mis en œuvre.

4.1.1. Indexation : chaînes thématique, temporelle et spatiale dédiées

Conformément aux préconisations de Clough *et al.* (2006), nous traitons indépendamment chacune des trois dimensions de l'information géographique : spatiale,

temporelle et thématique. Ceci implique la construction de plusieurs index, un par dimension au moins, comme le préconise (Martins *et al.*, 2005). La recherche monodimensionnelle et la gestion des index (ajout de nouveaux documents dans le corpus) restent efficaces. Cette approche vise également la combinaison de ces index dans le cadre de recherches d'information multidimensionnelle. Elle contribue au domaine de la RI géographique telle que défini par Jones *et al.* (2006). La figure 1 décrit les trois chaînes d'indexation dédiées au traitement de documents textuels dans le prototype PIV.

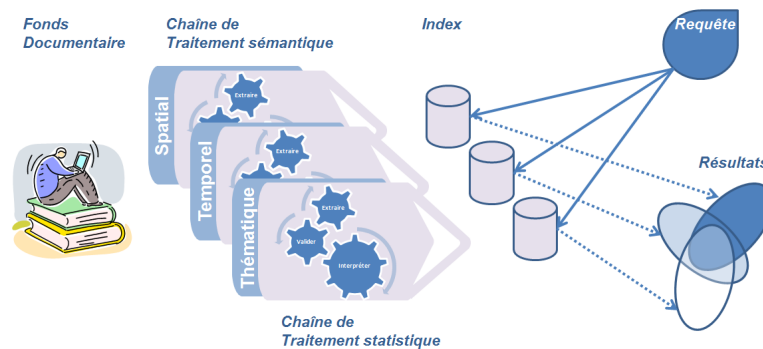


Figure 1. Les chaînes de traitement de PIV dédiées aux corpus de documents textuels.

Chaque chaîne de traitement génère un index spécialisé. Les chaînes spatiale et temporelle sont supportées par des modules de traitement automatique de la langue. Elles conduisent à l'extraction et à l'interprétation d'entités spatiales (ES) et temporelles (ET) contenues dans des documents textuels : « *le gave de Pau* » est annoté ES absolue (ESA) tandis que « *au nord du gave de Pau* » est annoté ES relative (ESR) – relation spatiale d'orientation (Gaio *et al.*, 2008) ; de même « *le printemps 1840* » est annoté ET absolue (ETA) tandis que « *vers le printemps 1840* » est annoté ET relative (ETR) – relation temporelle d'adjacence (Le Parc-Lacayrelle *et al.*, 2007).

Ainsi, la création des index spatial et temporel est réalisée en quatre principales étapes (figure 2). La première étape consiste à extraire les ES et les ET à l'aide d'une chaîne de traitement lexico-syntaxique (Gaio *et al.*, 2008). Cette chaîne est supportée par la plateforme Linguastream (Bilhaut *et al.*, 2003; Widlöcher *et al.*, 2005). Après un prétraitement de segmentation du contenu (*tokenization*) classique, nous adoptons un comportement de lecture active : un marqueur de *tokens* candidats localise ceux correspondant à des entités spatiales ou temporelles en utilisant des règles typographiques et lexicales. Une analyse morpho-syntaxique reconstitue ensuite des groupes nominaux correspondant à ces ES et ET candidates (« *gave de Pau* », par exemple).

La deuxième étape détermine une représentation symbolique pour chacune des ES et ET candidates détectées précédemment. Des règles de grammaire DCG (*Definite Clause Grammar*) permettent l'analyse sémantique de chaque ES et de chaque ET récursivement afin de déterminer s'il s'agit d'une ESA, ESR, ETA ou ETR. Ainsi, lorsque des relations spatiales ou temporelles sont détectées au sein d'un groupe no-

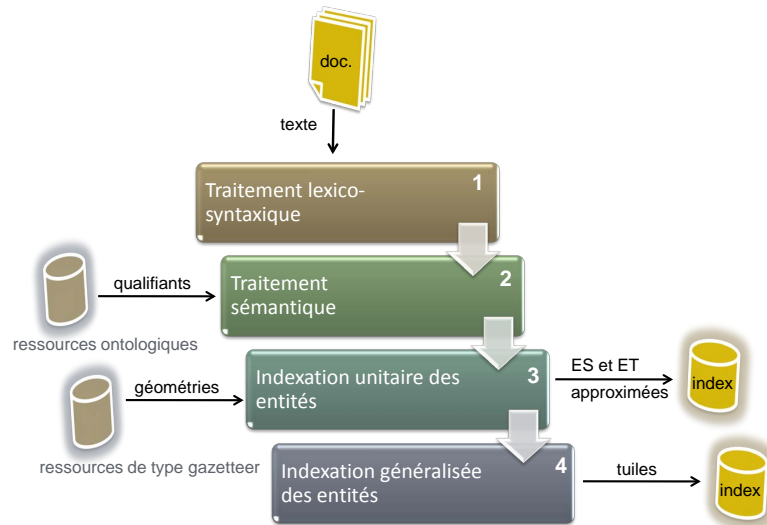


Figure 2. Les quatre principales étapes d'indexation dans le prototype PIV.

minal candidat, chaque relation est analysée et des étiquettes (telles que relation d'adjacence, d'inclusion, d'orientation, de distance, d'union) sont associées à l'ESR ou l'ETR correspondante. De même, une analyse de chaque ES et ET supportée par des règles de grammaire DCG dédiées au typage de ces groupes nominaux candidats permet d'apposer une seconde catégorie d'étiquettes (oronyme, hydronyme, voie de communication, commune, lieu-dit, etc.) qualifiant ces entités. Ainsi, par exemple, le syntagme nominal « au nord du gave de Pau » est annoté ESR avec une première étiquette *relation spatiale d'orientation – nord* et une seconde étiquette *qualifiant – hydronyme*. Cette ESR est définie récursivement à partir de l'ESA de toponyme *Pau*. L'ontologie géographique développée dans le cadre du projet ANR GEONTO (Kergosien *et al.*, 2009) est utilisée pour déterminer ces qualifiants.

La troisième étape interprète les représentations symboliques ainsi obtenues et calcule des représentations numériques. L'interprétation est supportée par des algorithmes d'approximation qui associent des intervalles de temps aux ET (Le Parc-Lacayrelle *et al.*, 2007) et des géométries aux ES (Gaio *et al.*, 2008), à l'aide des opérateurs spatiaux du SIG PostGIS³. Concernant l'approximation spatiale, nous utilisons des ressources de type gazetteer (base BD NYME[®] de l'IGN, ressources contributives locales, etc.) pour valider et géolocaliser chaque ESA. Les relations spatiales (par exemple, orientation – nord) sont ensuite interprétées et leurs représentations calculées ; des opérateurs SIG (par exemple, translation, intersection) sont appliqués à l'ES de référence. Pour l'approximation temporelle, selon le même principe, un in-

3. <http://postgis.refractions.net>

tervalle ou une date est d'abord associé à chaque ETA, puis, les relations temporelles (par exemple, adjacence – vers) sont interprétées et un ou plusieurs intervalles sont calculés à partir de l'ET de référence. Un index spatial et un index temporel résultent de cette troisième étape. L'index spatial décrit chaque ES avec la géométrie, le syntagme, l'identifiant du paragraphe et du document correspondant. De même, l'index temporel décrit chaque ET avec la période, le syntagme, l'identifiant du paragraphe et du document correspondant. Ce premier niveau d'index supporte plusieurs scénarios de recherche d'information : le recouvrement des zones spatiales ou temporelles de l'index et de la requête peut être calculé et l'ensemble des paragraphes pertinents retourné (Le Parc-Lacayrelle *et al.*, 2007; Sallaberry *et al.*, 2008).

Enfin, la quatrième étape uniformise les index ainsi constitués. Il s'agit d'un tui-lage spatial, temporel et thématique qui permet notamment de mémoriser les fré-quences d'évocation de tout ou partie de ces tuiles spatiales et temporelles dans les textes (Palacio *et al.*, 2009; Palacio *et al.*, 2010). Cette approche nécessite des traite-ments complémentaires liés aux trois dimensions :

1) spatiale : permettant un carroyage régulier ou administratif (commune, canton, département, etc.) de la zone couverte par le fonds documentaire et une projection des entités spatiales de l'index (et de leur fréquence d'apparition) dans ce carroyage ;

2) temporelle : permettant un carroyage régulier ou calendaire (jour, semaine, mois, etc.) de la période couverte par le fonds documentaire et une projection des entités temporelles de l'index (et de leur fréquence d'apparition) dans ce carroyage ;

3) thématique : permettant un carroyage (basé sur les concepts issus d'ontologies de domaine spécifique) des sujets couverts par le fonds documentaire et une projection des termes de l'index (et de leur fréquence d'apparition) dans ce carroyage.

L'intérêt de cette étape d'uniformisation est double. Elle permet tout d'abord de ramener des représentations spatiales, temporelles et thématiques diverses à une représentation homogène supportée par un redécoupage uniforme de l'espace, du temps et du thème. Elle permet également la mise en œuvre de stratégies de RI et de calculs de scores de pertinence éprouvés basés sur les fréquences d'apparition de ces carreaux spatiaux, temporels ou thématiques dans les unités documentaires. Une telle généra-lisation, c'est-à-dire l'uniformisation conduisant au regroupement d'entités géogra-phiques dans des tuiles, entraîne une perte de précision (Chrisman, 1990). Toutefois, elle est compensée par l'introduction de calculs de fréquences ainsi que la produc-tion et l'utilisation d'index de grains différents. Ainsi, par exemple, l'index communal (respectivement cantonal) sera utilisé par le moteur de recherche si la requête vise une ville particulière (respectivement une zone plus large). Ces aspects sont repris dans la section suivante qui traite de l'interrogation des index.

4.1.2. *Interrogation : combinaison des listes de résultats*

Dans le cadre de PIV, une requête géographique est exprimée sous forme de texte libre, comme présenté dans la section 1. Les dimensions concernées et le grain de précision requis sont identifiés dans cette expression. Puis, seuls les index correspon-

dants sont interrogés. De ce fait, qu’elles portent sur une, deux ou trois dimensions, les requêtes sont traitées uniformément.

Parmi les options d’indexation possibles et d’après les résultats expérimentaux présentés dans (Palacio *et al.*, 2010), nous avons retenu un carroyage administratif communal (respectivement mensuel) pour l’indexation spatiale (respectivement temporelle), dans le cadre de ce travail d’expérimentation. À ces carroyages, nous avons appliqué des calculs de fréquences discrets et continus (proportionnels à la surface de recouvrement entre l’ES et la tuile spatiale, ou bien entre l’ET et la tuile temporelle). Pour la partie RI, nous avons préalablement expérimenté différents modèles de RI (TF, TF-IDF, OkapiBM25) que nous avons appliqués à des pondérations discrètes et continues. Les résultats de notre étude nous ont permis de retenir la formule du TF associée à une pondération continue (Palacio *et al.*, 2009; Palacio *et al.*, 2010). La chaîne de traitement thématique n’étant pas encore totalement automatisée, nous proposons d’évaluer nos travaux relatifs à l’espace et au temps et de limiter la partie thématique au système d’indexation plein-texte Terrier (Ounis *et al.*, 2005).

Ainsi, chaque SRI monodimensionnel est indépendant : il construit et interroge un index qui lui est propre. Dans notre approche, le SRI géographique proposé repose sur plusieurs SRI monodimensionnels (sources) dont les résultats sont combinés pour ne former qu’une seule liste de résultats l . Or, dans la littérature de RI, Fox *et al.* (1993) ont proposé à cet effet le combinateur CombMNZ. La liste combinée l comprend alors tous les documents distincts restitués par les SRI sources. Dans cette liste, la similarité s d’un document d est calculée en additionnant les similarités de d issues des sources, cette somme étant pondérée par le nombre de SRI sources ayant restitué d . Ainsi, pour une requête q donnée, d sera d’autant plus pertinent dans l (c’est-à-dire classé en tête de l) qu’il a été restitué par de nombreux SRI en tête de liste (similarité s élevée entre d et q). Le comportement de CombMNZ est assimilable au principe du faisceau de preuves : des documents restitués par plusieurs SRI constituent autant d’indices renforçant la présomption de pertinence eu égard à ces documents. Pour un autre contexte combinant les dimensions thématique et sémantique, ce principe a précédemment été validé expérimentalement dans (Hubert *et al.*, 2009). De plus, Lee (1997) a comparé les résultats de CombMNZ avec d’autres opérateurs sur des collections de test de TREC, montrant son efficacité. C’est cette qualité avérée qui nous a motivés à utiliser cet opérateur pour combiner les résultats des SRI monodimensionnels. Les similarités s qu’ils calculent étant très variables d’un SRI à l’autre – appartenant à des intervalles réels différents – nous les normalisons sur $[0; 1]$ au préalable grâce à l’équation 1 proposée dans (Lee, 1997).

$$\text{similarité_normalisée} = \frac{\text{similarité_non_normalisée} - \text{similarité_minimum}}{\text{similarité_maximum} - \text{similarité_minimum}} \quad [1]$$

Ainsi, pour une requête donnée $q = 8$, la figure 3 (a-c) illustre les résultats des trois SRI, chacun étant composé de couples (d_i, s) où d_i est un document et s est la similarité calculée entre q et d_i . Le résultat de la combinaison des SRI est représenté en

figure 3 (d) où les valeurs des similarités résultant de CombMNZ sont présentées avec le détail des calculs associés, basés sur les similarités des sources (a–c) normalisées. Sur cet exemple, on observe le fait que le score d'un d_i dépend de deux facteurs. Plus un d_i est souvent restitué par les SRI, plus il obtient un score s élevé. En complément, ce dernier est d'autant plus élevé que le d_i était initialement classé en tête de liste. Le document d_4 illustre ce principe.

(a) SRI thématique	(b) SRI spatial	(c) SRI temporel																																																			
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">q</th> <th style="border: none;">d</th> <th style="border: none;">s</th> </tr> </thead> <tbody> <tr><td style="border: none;">8</td><td style="border: none;">d_4</td><td style="border: none;">14,5</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_3</td><td style="border: none;">12</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_7</td><td style="border: none;">8,7</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_1</td><td style="border: none;">0,5</td></tr> </tbody> </table>	q	d	s	8	d_4	14,5	8	d_3	12	8	d_7	8,7	8	d_1	0,5	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">q</th> <th style="border: none;">d</th> <th style="border: none;">s</th> </tr> </thead> <tbody> <tr><td style="border: none;">8</td><td style="border: none;">d_8</td><td style="border: none;">150</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_1</td><td style="border: none;">120</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_4</td><td style="border: none;">80</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_9</td><td style="border: none;">-10</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_2</td><td style="border: none;">-30</td></tr> </tbody> </table>	q	d	s	8	d_8	150	8	d_1	120	8	d_4	80	8	d_9	-10	8	d_2	-30	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">q</th> <th style="border: none;">d</th> <th style="border: none;">s</th> </tr> </thead> <tbody> <tr><td style="border: none;">8</td><td style="border: none;">d_8</td><td style="border: none;">1</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_4</td><td style="border: none;">0,7</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_9</td><td style="border: none;">0,5</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_1</td><td style="border: none;">0,5</td></tr> <tr><td style="border: none;">8</td><td style="border: none;">d_2</td><td style="border: none;">0,5</td></tr> </tbody> </table>	q	d	s	8	d_8	1	8	d_4	0,7	8	d_9	0,5	8	d_1	0,5	8	d_2	0,5
q	d	s																																																			
8	d_4	14,5																																																			
8	d_3	12																																																			
8	d_7	8,7																																																			
8	d_1	0,5																																																			
q	d	s																																																			
8	d_8	150																																																			
8	d_1	120																																																			
8	d_4	80																																																			
8	d_9	-10																																																			
8	d_2	-30																																																			
q	d	s																																																			
8	d_8	1																																																			
8	d_4	0,7																																																			
8	d_9	0,5																																																			
8	d_1	0,5																																																			
8	d_2	0,5																																																			
↘	↓	↙																																																			
(d) SRI géographique résultant de la combinaison des trois résultats																																																					
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">q</th> <th style="border: none;">d</th> <th style="border: none;">Similarité s calculée avec CombMNZ</th> </tr> </thead> <tbody> <tr> <td style="border: none;">8</td> <td style="border: none;">d_4</td> <td style="border: none;">$6,0333 = 3 \times \left(\frac{14,5-0,5}{14,5-0,5} + \frac{80+30}{150+30} + \frac{0,7-0,5}{1-0,5} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_8</td> <td style="border: none;">$4,0000 = 2 \times \left(\frac{150+30}{150+30} + \frac{1-0,5}{1-0,5} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_1</td> <td style="border: none;">$2,5000 = 2 \times \left(\frac{0,5-0,5}{14,5-0,5} + \frac{120+30}{150+30} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_3</td> <td style="border: none;">$0,8214 = 1 \times \left(\frac{12-0,5}{14,5-0,5} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_7</td> <td style="border: none;">$0,5857 = 1 \times \left(\frac{8,7-0,5}{14,5-0,5} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_9</td> <td style="border: none;">$0,2222 = 2 \times \left(\frac{-10+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$</td> </tr> <tr> <td style="border: none;">8</td> <td style="border: none;">d_2</td> <td style="border: none;">$0,0000 = 2 \times \left(\frac{-30+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$</td> </tr> </tbody> </table>			q	d	Similarité s calculée avec CombMNZ	8	d_4	$6,0333 = 3 \times \left(\frac{14,5-0,5}{14,5-0,5} + \frac{80+30}{150+30} + \frac{0,7-0,5}{1-0,5} \right)$	8	d_8	$4,0000 = 2 \times \left(\frac{150+30}{150+30} + \frac{1-0,5}{1-0,5} \right)$	8	d_1	$2,5000 = 2 \times \left(\frac{0,5-0,5}{14,5-0,5} + \frac{120+30}{150+30} \right)$	8	d_3	$0,8214 = 1 \times \left(\frac{12-0,5}{14,5-0,5} \right)$	8	d_7	$0,5857 = 1 \times \left(\frac{8,7-0,5}{14,5-0,5} \right)$	8	d_9	$0,2222 = 2 \times \left(\frac{-10+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$	8	d_2	$0,0000 = 2 \times \left(\frac{-30+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$																											
q	d	Similarité s calculée avec CombMNZ																																																			
8	d_4	$6,0333 = 3 \times \left(\frac{14,5-0,5}{14,5-0,5} + \frac{80+30}{150+30} + \frac{0,7-0,5}{1-0,5} \right)$																																																			
8	d_8	$4,0000 = 2 \times \left(\frac{150+30}{150+30} + \frac{1-0,5}{1-0,5} \right)$																																																			
8	d_1	$2,5000 = 2 \times \left(\frac{0,5-0,5}{14,5-0,5} + \frac{120+30}{150+30} \right)$																																																			
8	d_3	$0,8214 = 1 \times \left(\frac{12-0,5}{14,5-0,5} \right)$																																																			
8	d_7	$0,5857 = 1 \times \left(\frac{8,7-0,5}{14,5-0,5} \right)$																																																			
8	d_9	$0,2222 = 2 \times \left(\frac{-10+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$																																																			
8	d_2	$0,0000 = 2 \times \left(\frac{-30+30}{150+30} + \frac{0,5-0,5}{1-0,5} \right)$																																																			

Figure 3. Principe de combinaison de résultats de recherche avec CombMNZ.

4.2. Mise en œuvre du cadre expérimental pour évaluer le SRI géographique PIV

Afin d'évaluer le SRI géographique PIV, nous avons mis en œuvre le cadre expérimental présenté en section 3. Nous détaillons successivement la collection de test constituée, la validation du choix du combinateur CombMNZ, les analyses comparatives réalisées (niveau global, puis par topic) ainsi que les limites identifiées.

4.2.1. Constitution de la collection de test MIDR_2010

La collection de test MIDR_2010⁴ comprend les quatre volets identifiés en section 3.1. Premièrement, le corpus documentaire représente 5 645 paragraphes issus

4. Médiathèque Intercommunale à Dimension Régionale, située à Pau en Aquitaine. La collection de test est accessible sur <http://t2i.univ-pau.fr/MIDR>.

de 11 ouvrages numérisés (et traités avec une application de reconnaissance de caractères) qui proviennent du fonds patrimonial de la médiathèque. Un document d restitué à l'utilisateur par le SRI géographique est un de ces paragraphes, vu comme le meilleur point d'entrée dans l'ouvrage associé. Deuxièmement, 33 *topics* couvrant tout ou partie des trois dimensions de l'information géographique ont été constitués. Troisièmement, les *qrels* ont été obtenus en interrogeant trois SRI – un thématique basé sur le modèle PL2 (configuration de base de Terrier), un spatial et un temporel – avec le titre des *topics*. Quatrièmement, les ressources géographiques liées au corpus sont issues de l'Institut Géographique National (BD NYME[®]) et d'un *gazetteer* (un dictionnaire géographique) contributif local, ainsi que d'une base de connaissances calendaire. Pour chaque *topic*, les résultats restitués par tous les SRI ont été pris en compte pour constituer le pool. Ce dernier a été évalué, en considérant un jugement booléen par dimension ainsi qu'un jugement booléen global, ces quatre évaluations étant agrégées pour constituer un jugement graduel, comme expliqué en section 3.1.

4.2.2. Validation du choix du combineur CombMNZ normalisé

Comme indiqué dans la section 4.1.2, PIV combine les résultats des trois SRI monodimensionnels à l'aide du combineur CombMNZ normalisé. Cette configuration est la plus performante parmi celles testées par Lee (1997). Dans cette section, nous vérifions qu'il en est de même pour la collection MIDR_2010, en confrontant les performances respectives des cinq combineurs détaillés dans (Lee, 1997), dont nous reprenons les définitions dans le tableau 1.

Combineur	Définition
CombMIN	Minimum des similarités individuelles
CombMAX	Maximum des similarités individuelles
CombSUM	Somme des similarités individuelles
CombANZ	$\text{CombSUM} \div \text{Nombre de similarités non nulles}$
CombMNZ	$\text{CombSUM} \times \text{Nombre de similarités non nulles}$

Tableau 1. Définition des combineurs (Fox et al., 1993; Lee, 1997).

Le tableau 2 montre les performances des combineurs appliqués aux trois dimensions (thématique, spatiale et temporelle) selon la mesure *NDCG*. Ces derniers sont déclinés en version normalisée et non normalisée. L'effet de la normalisation, inférieur à 5 %, est également indiqué. Cette expérimentation menée sur la collection MIDR_2010 corrobore les observations de Lee (1997) : elle montre que la configuration optimale (0,8029) correspond à CombMNZ normalisé.

Dans la section suivante, nous testons la performance de la combinaison basée sur CombMNZ normalisé, en fonction des dimensions utilisées.

4.2.3. Analyse comparative de la performance de toutes les combinaisons possibles

Le tableau 3 présente les comparaisons effectuées entre les différents SRI et deux « baselines » thématiques (SRI de référence) identifiées dans (Perea-Ortega et al.,

	Combinateurs Comb*				
	Min	Max	Sum	Anz	Mnz
Non normalisée	0,6141	0,7610	0,7873	0,7480	0,7946
Normalisée	0,6434	0,7624	0,7876	0,7400	0,8029
Gain (%)	4,77	0,18	0,04	-1,07	1,04

Tableau 2. Performances relatives de combineurs et effet de la normalisation.

2008) : Th^+ est une baseline forte correspondant au modèle OkapiBM25 et Th^- est une baseline faible correspondant au modèle TF-IDF. Le SRI spatial est noté S et le SRI temporel est noté Te (cf. section 3.2). Calculés avec le programme `trec_eval`⁵ version 9.0 utilisé à TREC, ces résultats portent sur l'analyse des performances des moteurs au regard des 33 requêtes.

Fusion de N SRI	SRI monodimensionnels				moyNDCG	Amélioration (%)	
	Th^-	Th^+	S	Te		Th^-	Th^+
1	✓				0,4810	0,0	0,0
		✓			0,4809	0,0	0,0
			✓		0,4956	3,0	3,1
				✓	0,4875	1,4	1,4
2	✓	✓			0,4809	0,0	0,0
	✓		✓		0,6444*†	34,0	34,0
	✓			✓	0,7081*†	47,2	47,2
		✓	✓		0,6451*†	34,1	34,1
		✓		✓	0,7086*†	47,3	47,3
3			✓		0,7262*†	51,0	51,0
	✓	✓	✓		0,6375*†	32,5	32,6
	✓	✓		✓	0,6909*†	43,6	43,7
	✓		✓	✓	0,8013*†	66,6	66,6
4		✓	✓	✓	0,8029*†	66,9	67,0
	✓	✓	✓	✓	0,7732*†	60,7	60,8

Tableau 3. Efficacité des SRI par rapport aux baselines thématiques. Le symbole ‘*’ (resp. †) indique une différence significative par rapport à la baseline Th^- (resp. Th^+).

Globalement, il est intéressant de noter que ces deux baselines sont quasiment identiques. Contrairement aux résultats rapportés par Perea-Ortega *et al.* (2008), TF-IDF fournit de meilleurs résultats qu'OkapiBM25 dans notre cas. Cette différence peut être due au fait que la collection MIDR_2010 est décomposée en paragraphes de document de tailles homogènes (*versus* des documents entiers de tailles variables). Or, la bonne performance d'OkapiBM25 s'observe notamment sur un corpus aux documents de tailles variables, ce qui n'est pas le cas dans le présent corpus.

5. http://trec.nist.gov/trec_eval

Concernant les moteurs monodimensionnels, on observe une performance maximale de 0,4956 pour le SRI spatial. Ces SRI sont toutefois caractérisés par des performances similaires. Par ailleurs, le gain apporté par la combinaison de dimensions hétérogènes est observable à partir de deux dimensions combinées. Notons que ce gain est statistiquement significatif au regard des deux baselines. Nous remarquons que l’alliance du spatial et du temporel fournit les meilleures performances (0,7262). Toutefois, la combinaison du thématique et du temporel offre une performance similaire (0,7086). Ce résultat est certainement dû au fait qu’une entité spatiale absolue, telle que « *Gavarnie* », est détectable par un SRI thématique. Cependant, des situations plus complexes faisant intervenir des entités spatiales relatives ne pourront être correctement traitées qu’avec un SRI spatial.

Enfin, la combinaison des trois dimensions apporte la performance maximale (0,8029). Le gain de précision correspondant à 67,0 % par rapport à la baseline forte Th^+ valide l’hypothèse à la base de notre travail : combiner les trois dimensions de l’information géographique améliore la pertinence des résultats par rapport à la dimension thématique seule. En complément, une analyse approfondie montre que combiner les trois dimensions offre des résultats significativement supérieurs (13,3 %, $p < 0,001$) à la meilleure combinaison de deux dimensions (S+Te). Notons qu’une combinaison des trois dimensions avec les deux versions thématiques (Th^+ et Th^-) n’apporte pas d’amélioration supplémentaire (0,7732), le renforcement thématique masquant alors les aspects complémentaires apportés par les deux autres dimensions.

4.2.4. Analyse par topic du SRI tridimensionnel PIV

Après avoir identifié la combinaison de dimensions optimisant la performance de PIV (Th^+ +S+Te), nous analysons dans cette section sa performance par topic. Cette analyse plus détaillée a pour objectif d’affiner les observations précédentes. À cet effet, le tableau 4 synthétise, pour chaque topic :

- 1) le numéro du topic (colonne 1),
- 2) le nombre de documents jugés pertinents (cf. *qrels*) pour chacune des trois dimensions (colonnes 2 à 4),
- 3) la performance de PIV pour le topic considéré (colonne 5),
- 4) et le nombre de documents pertinents restitués par un seul, par uniquement deux ou par les trois SRI, chacun étant associé à une dimension (colonnes 6 à 8).

La répartition des documents pertinents par dimension est variable d’un topic à l’autre. On remarque une prédominance globale de la dimension spatiale. Ceci est dû à la spécificité des documents de la collection MIDR_2010 qui traitent de sujets liés à la zone spatiale correspondant aux Pyrénées. Par ailleurs, l’écart-type faible $\sigma_{NDCG} = 0,0702$ indique une performance par topic stable. Globalement, la différence de performance par rapport à la moyenne de 0,8029 calculée sur les 33 topics est dans $[-20\% ; +13\%]$. Enfin, il est intéressant d’étudier la contribution des trois dimensions : sont-elles redondantes (restituant des documents similaires) ou bien complémentaires (restituant des documents spécifiques à chaque dimension) ? La dernière

Topic	Nombre de documents pertinents			NDCG	Complémentarité		
	Th ⁺	S	Te		1 dim	2 dim	3 dim
1	67	260	134	0,7707	148	134	15
2	60	213	138	0,8758	195	87	14
3	42	111	101	0,7906	201	25	1
4	16	16	74	0,5978	82	9	2
5	94	87	34	0,8413	142	35	1
6	48	60	59	0,8494	139	14	1
7	71	24	93	0,7348	127	29	1
8	27	177	103	0,8097	229	39	1
9	41	209	112	0,7949	154	95	6
10	68	110	102	0,7703	176	52	6
11	33	111	6	0,7105	90	30	6
12	41	246	116	0,7523	150	101	17
13	106	120	47	0,6643	201	33	2
14	74	110	107	0,7857	137	74	2
15	89	152	103	0,7010	190	74	2
16	74	40	0	0,7899	84	15	2
17	197	116	23	0,7161	190	70	2
18	37	141	112	0,8757	218	36	2
19	113	175	101	0,7162	199	83	8
20	180	11	73	0,7332	147	57	1
21	13	150	122	0,7068	157	58	4
22	7	162	121	0,8873	198	46	4
23	126	114	120	0,8356	205	58	13
24	170	71	30	0,8634	184	42	1
25	234	102	110	0,8155	132	100	38
26	160	227	131	0,7902	71	126	65
27	19	41	113	0,8507	149	12	65
28	21	94	117	0,8323	139	45	1
29	47	41	120	0,7748	149	25	3
30	51	91	131	0,8758	213	30	3
31	16	194	55	0,8487	185	40	3
32	5	224	9	0,7671	160	93	2
33	23	60	55	0,7552	84	4	2
Moyenne	72	123	91	0,8029	158	54	9

Tableau 4. *Étude par topic de la distribution des documents pertinents selon les trois dimensions, de la performance du SRI PIV et de la complémentarité des dimensions.*

colonne du tableau 4 montre un faible nombre de documents simultanément restitués par les trois dimensions, démontrant par conséquent que les trois dimensions sont complémentaires.

4.2.5. Limites actuelles de l'évaluation

En l'état, l'expérimentation proposée dans cet article présente au moins deux limites. Premièrement, de par ses 5 645 paragraphes totalisant 3,7 Mo, la collection de test MIDR_2010 est très peu volumineuse en comparaison des collections TREC. Deuxièmement, nous avons réalisé les analyses à partir de 33 topics. Cette valeur représente huit topics de plus que le nombre minimum de topics à considérer pour pouvoir réaliser des analyses statistiques valides. Nous continuons l'effort de jugement manuel des documents permettant d'obtenir davantage de topics à analyser.

Malgré ces limites, le cadre d'évaluation présenté dans cet article est applicable aux diverses propositions issues du domaine de la RI géographique. La section suivante présente de façon succincte un sous-ensemble représentatif de ces travaux.

5. Approches de la littérature en RI géographique

Des travaux connexes au projet PIV sont menés dans la communauté internationale; nous pouvons notamment citer les sept projets suivants. Le projet CITER "*Creation of a European History Textbook Repository*" (Pfoser *et al.*, 2009) permet d'effectuer des recherches sur des livres historiques via plusieurs dimensions. Le projet DIGMAP "*Discovering our Past World with Digitised Maps*" (Manguinhas *et al.*, 2009) est un système spécialement axé sur le matériel historique et la valorisation du patrimoine culturel et scientifique. Le projet GEOSEM (Bilhaut *et al.*, 2003) est un système dédié au traitement sémantique pour l'information géographique (textes, cartes, graphiques). Le projet GIPSY "*Georeferenced Information Processing System*" (Woodruff *et al.*, 1994) propose une méthode d'indexation de documents textuels basée sur l'agrégation des géo-références correspondant aux entités spatiales trouvées dans le texte. L'idée est d'utiliser cette agrégation pour retrouver la zone géographique la plus représentative, qui servira à indexer le document. Le projet GRID "*Geospatial Retrieval of Indexed Documents*" (Valcartier, 2006) est un système dédié à la recherche d'information textuelle combinant une recherche par mots-clés et une recherche par zone d'intérêt via une interface cartographique. Le projet SPIRIT "*Spatially-Aware Information Retrieval on the Internet*" (Vaid *et al.*, 2005) est un système dédié à la recherche de pages web faisant référence à des lieux ou zones géographiques spécifiés dans une requête. Enfin, le projet STEWARD "*Spatio-Textual Extraction on the Web Aiding Retrieval of Documents*" (Lieberman *et al.*, 2007) est un système dédié à l'extraction, l'interrogation et la visualisation de références à des zones géographiques dans des textes non structurés. Comme indiqué dans le tableau 5, ces systèmes d'indexation et de RI géographique traitent en priorité la composante spatiale (entités ESA et ESR, cf. section 4.1.1).

Concernant la combinaison en phase d'interrogation, les systèmes CITER, GRID et SPIRIT adoptent une approche de « filtrage parallèle ». Ils interrogent simultanément et séparément chaque dimension, puis combinent les différentes listes de résultats en réalisant leur intersection. De par le résultat ensembliste de cette opération, la perti-

Système	Nature des index				Thématique
	Spatial		Temporel		
	ESA	ESR	ETA	ETR	
CITER	✓		✓		✓
DIGMAP	✓		✓		✓
GEOSEM	✓	✓	✓	✓	✓
GIPSY	✓				
GRID	✓				✓
PIV	✓	✓	✓	✓	✓
SPIRIT	✓				✓
STEWARD	✓				✓

Tableau 5. Comparaison de projets et prototypes dédiés à la RI géographique.

nence d'un document est booléenne, ce qui ne permet pas de classer les pertinents. D'autres systèmes, tel que STEWARD, mettent en œuvre une approche de « filtrage séquentiel » qui consiste à interroger le corpus sur une dimension puis à appliquer les autres dimensions sur le sous-ensemble de documents obtenu (priorité aux termes, puis au spatial). Ici les résultats sont classés selon les scores du dernier critère, donc pour STEWARD sur le spatial. Ces approches de filtrage permettent de combiner différents critères sans les uniformiser. Néanmoins, cela ne permet pas de calculer un score global et donc de les classer selon les différentes dimensions impliquées. D'autres systèmes, tels que DIGMAP et GEOSEM, proposent d'utiliser des approches de combinaison linéaire (par exemple la moyenne arithmétique) sans toutefois uniformiser les critères, ce qui peut biaiser le score global.

Nous avons présenté, dans le tableau 5, les principales caractéristiques des index générés dans ces prototypes. Cette étude doit être complétée par l'évaluation de la performance de ces systèmes au regard de la pertinence de leurs résultats. Notons que le cadre proposé est adapté à cette tâche, permettant par ailleurs la comparaison des SRI aussi divers soient-ils, comme constaté dans le tableau 5.

6. Conclusion et perspectives

Dans cet article, nous avons considéré les SRI géographique exploitant les dimensions spatiale, temporelle et thématique. Or, les moteurs de recherche usuels s'avèrent limités dans de tels contextes. Aussi, notre contribution se décline en trois volets : 1) la proposition d'un cadre d'évaluation pour la RIG ainsi que 2) la validation de notre hypothèse initiale (combinaison des trois dimensions améliore la qualité des résultats de recherche). L'application de ce cadre sur une collection de test appropriée a mis en exergue une amélioration de 67,0 % qui est, de plus, statistiquement significative. L'étude approfondie, par topic, des résultats montre que 3) les trois dimensions ne sont pas redondantes, mais bien complémentaires. L'ensemble de ces résultats apporte une

validation empirique de nos propositions expérimentées avec le prototype PIV (Gao *et al.*, 2008).

Ce cadre nous a non seulement permis d'évaluer notre approche, mais aussi d'en identifier certains défauts. Par exemple, les chaînes de traitement spatiale et temporelle restituait des documents avec une similarité très faible dans le domaine $[0; 1]$, tronquée de surcroît à quatre positions décimales. De nombreux documents étaient alors *ex aequo* dans les listes de résultats. Dans ce cas, *trec_eval* utilise le nom des documents pour les départager, introduisant ainsi un biais expérimental : la performance des systèmes dépend de leur qualité intrinsèque ainsi que d'un facteur chance lié au nommage des documents. Ce « biais des *ex aequo* » et la quantification de son effet sur plusieurs campagnes de TREC sont détaillés dans (Cabanac *et al.*, 2010a; Cabanac *et al.*, 2010b). Nous avons corrigé ce défaut en veillant à ce que les chaînes de traitement fournissent des résultats davantage discriminés.

Par ailleurs, le cadre permet également et dès à présent :

1) de comparer des variantes pour chacune des trois dimensions (par exemple, différents types de tuilage pour la dimension spatiale). En complément, est également possible l'intégration d'autres dimensions, telles que la confiance en l'information et sa fraîcheur (Costa Pereira *et al.*, 2009) ;

2) de tester et comparer de nouveaux opérateurs de combinaison. Il est ainsi possible de confronter la performance des combineurs basés sur les valeurs de similarité des documents (de type Comb* (Fox *et al.*, 1993)) avec celles des combineurs basés sur les rangs des documents, comme celui défini dans (de Borda, 1781), par exemple ;

3) de tester l'influence de la pondération des dimensions au sein de PIV, pour donner une préférence à la dimension thématique, par exemple. Ceci nous permettrait d'ajuster notre modèle pour le rapprocher de la perception humaine, basée sur les priorités établies mentalement entre dimensions ;

4) de comparer notre SRI PIV avec ceux identifiés dans la section 5, notamment.

En termes de perspectives, nous nous intéresserons plus particulièrement aux approches de combinaison par contraintes (prenant en compte les notions d'exigences et de préférences) selon une approche linéaire (Farah *et al.*, 2008), harmonique (?) ou selon une approche basée sur la logique floue avec les opérateurs OWA (Yager, 1988; Boughanem *et al.*, 2007), ou encore, sur d'autres propositions comme l'intégrale de Choquet (Labreuche *et al.*, 2003).

Ayant montré dans cet article la faisabilité d'une évaluation de SRI géographique, nous envisageons également de proposer cette tâche pour une campagne de type GeoCLEF, son originalité étant de traiter des documents sur les trois dimensions : thématique, spatiale et temporelle. La collection de test MIDR_2010 constituée est d'ores et déjà disponible sur le site web du projet PIV⁶. À plus long terme, il s'agira de compléter la collection de test, en termes de documents comme de topics. L'aspect

6. <http://t2i.univ-pau.fr/MIDR>

multilingue pourra également être intégré pour permettre la comparaison des SRI qui sont principalement axés sur le traitement de documents en anglais.

Remerciements

Les auteurs remercient vivement Valérie Bougeant, Jean-Paul Cabanac, Christian Chevalier, Benoît Chirle, Firas Damak, Agnès Itier, Rudy Kapler, Anaïs Lefevre, Nicolas Noullet et Camille Raymond pour leur contribution bénévole en qualité d'as-seisseurs.

7. Bibliographie

- Bilhaut F., Charnois T., Enjalbert P., Mathet Y., "Geographic reference analysis for geographic document querying", *HLT-NAACL'03: Proceedings of the workshop on Analysis of geographic references*, ACL, Morristown, NJ, USA, p. 55–62, 2003.
- Boughanem M., Loiseau Y., Prade H., "Refining Aggregation Functions for Improving Document Ranking in Information Retrieval", *SUM'07: Proceedings of the 1st international conference on Scalable Uncertainty Management*, Springer-Verlag, Berlin, Heidelberg, p. 255–267, 2007.
- Bucher B., Clough P., Joho H., Purves R., Syed A. K., "Geographic IR Systems: Requirements and Evaluation", *ICC'05: Proceedings of the 22nd International Cartographic Conference*, Global Congressos, 2005. CDROM.
- Buckley C., Voorhees E. M., "Evaluating Evaluation Measure Stability", *SIGIR'00: Proceedings of the 23rd international ACM SIGIR conference*, ACM, New York, NY, USA, p. 33–40, 2000.
- Cabanac G., Hubert G., Boughanem M., Chrisment C., "Impact du « biais des *ex aequo* » dans les évaluations de Recherche d'Information", *CORIA'10 : Actes de la 7^e conférence en recherche d'information et applications*, p. 83–98, March, 2010a.
- Cabanac G., Hubert G., Boughanem M., Chrisment C., "Tie-breaking Bias: Effect of an Uncontrolled Parameter on Information Retrieval Evaluation", in M. Agosti, N. Ferro, C. Peters, M. de Rijke, A. F. Smeaton (eds), *CLEF'10 : Proceedings of the 1st Conference on Multilingual and Multimodal Information Access Evaluation*, vol. 6360 of *LNC3*, Springer-Verlag, p. 112–123, September, 2010b.
- Chrisman N. R., "Deficiencies of sheets and tiles: building sheetless databases", *Int. J. Geogr. Inf. Sci.*, vol. 4, n° 2, p. 157–167, 1990.
- Clough P., Joho H., Purves R., "Judging the Spatial Relevance of Documents for GIR", *ECIR'06: Proceedings of the 28th European Conference on IR Research*, vol. 3936 of *LNC3*, Springer, p. 548–552, 2006.
- Costa Pereira C., Dragoni M., Pasi G., "Multidimensional Relevance: A New Aggregation Criterion", *ECIR'09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, Springer-Verlag, Berlin, Heidelberg, p. 264–275, 2009.
- de Borda J.-C., "Mémoire sur les élections au scrutin", *Histoire de l'Académie Royale des Sciences*, Paris, 1781.

- Farah M., Vanderpooten D., “An outranking approach for information retrieval”, *Inf. Retr.*, vol. 11, n° 4, p. 315–334, 2008.
- Fox E. A., Shaw J. A., “Combination of Multiple Searches”, in D. K. Harman (ed.), *TREC-1: Proceedings of the First Text REtrieval Conference*, NIST, Gaithersburg, MD, USA, p. 243–252, February, 1993.
- Gaio M., *Traitements de l’information géographique : Représentations et structures*, Habilitation à diriger des recherches, Université de Caen, 2001.
- Gaio M., Sallaberry C., Etcheverry P., Marquesuzaa C., Lesbegueries J., “A global process to access documents’ contents from a geographical point of view”, *J. Vis. Lang. Comput.*, vol. 19, n° 1, p. 3–23, 2008.
- Gan Q., Attenberg J., Markowetz A., Suel T., “Analysis of geographic queries in a search engine log”, *LocWeb’08: Proceedings of the first international workshop on Location and the web*, ACM, New York, NY, USA, p. 49–56, 2008.
- Gey F. C., Larson R. R., Sanderson M., Joho H., Clough P., Petras V., “GeoCLEF’05: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview”, *CLEF’05: Proceedings of the 6th workshop on Cross-Language Evaluation Forum*, vol. 4022 of *LNCS*, Springer, p. 908–919, 2006.
- Gospodnetić O., Hatcher E., *Lucene in Action*, Manning Publications, 2005.
- Harman D. K., “The TREC Test Collections”, in Voorhees *et al.* (2005), chapter 2, p. 21–53, 2005.
- Hubert G., Mothe J., “An adaptable search engine for multimodal information retrieval”, *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, n° 8, p. 1625–1634, 2009.
- Hull D., “Using Statistical Testing in the Evaluation of Retrieval Experiments”, *SIGIR’93: Proceedings of the 16th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 329–338, 1993.
- Järvelin K., Kekäläinen J., “Cumulated gain-based evaluation of IR techniques”, *ACM Trans. Inf. Syst.*, vol. 20, n° 4, p. 422–446, 2002.
- Jones C. B., Purves R., “GIR’05 2005 ACM workshop on geographical information retrieval”, *SIGIR Forum*, vol. 40, n° 1, p. 34–37, 2006.
- Jones R., Zhang W. V., Rey B., Jhala P., Stipp E., “Geographic intention and modification in web search”, *Int. J. Geogr. Inf. Sci.*, vol. 22, n° 3, p. 229–246, 2008.
- Kanhabua N., Nørvåg K., “Improving Temporal Language Models for Determining Time of Non-timestamped Documents”, *ECDL’08: Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries*, Springer-Verlag, Berlin, Heidelberg, p. 358–370, 2008.
- Kergosien E., Kamel M., Sallaberry C., Bessagnet M.-N., Aussenac-Gilles N., Gaio M., “Construction automatique d’ontologie et enrichissement à partir de ressources externes”, *JFO’09: Actes des 3^e Journées Francophones sur les Ontologies*, France, p. 11–20, 2009.
- Labreuche C., Grabisch M., “The Choquet integral for the aggregation of interval scales in multicriteria decision making”, *Fuzzy Sets Syst.*, vol. 137, n° 1, p. 11–26, 2003.
- Le Parc-Lacayrelle A., Gaio M., Sallaberry C., “La composante temps dans l’information géographique textuelle”, *Document Numérique*, vol. 10, n° 2, p. 129–148, 2007.

- Lee J. H., "Analyses of Multiple Evidence Combination", *SIGIR'97: Proceedings of the 20th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 267–276, 1997.
- Lieberman M. D., Samet H., Sankaranarayanan J., Sperling J., "STEWART: Architecture of a Spatio-Textual Search Engine", *GIS'07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, ACM, New York, NY, USA, p. 1–8, 2007.
- Manguinhas H., Martins B., Borbinha J., Siabato W., "The DIGMAP Geo-Temporal Web Gazetteer Service", *e-Perimetron: Int. Web J. Sci. Technol. Affined Hist. Cartogr. Maps*, vol. 4, n° 1, p. 9–24, 2009.
- Manning C. D., Raghavan P., Schütze H., *Introduction to Information Retrieval*, Cambridge University Press, July, 2008.
- Martins B., Silva M. J., Andrade L., "Indexing and ranking in Geo-IR systems", *GIR'05: Proceedings of the 2005 workshop on Geographic information retrieval*, ACM, New York, NY, USA, p. 31–34, 2005.
- Ogilvie P., Callan J. P., "Experiments Using the Lemur Toolkit", *TREC'01: Proceedings of the 9th Text REtrieval Conference*, NIST, Gaithersburg, MD, USA, February, 2001.
- Ounis I., Amati G., Plachouras V., He B., Macdonald C., Johnson D., "Terrier Information Retrieval Platform", *ECIR'05: Proceedings of the 27th European Conference on IR Research*, vol. 3408 of LNCS, Springer, p. 517–519, 2005.
- Palacio D., Sallaberry C., Gaio M., "Normalizing Spatial Information to Better Combine Criteria in Geographical Information Retrieval", *ECIR-GIIW'09: Proceeding of the international workshop on Geographic Information on the Internet*, p. 37–48, 2009.
- Palacio D., Sallaberry C., Gaio M., "Normalizing Spatial Information to Improve Geographical Information Indexing and Retrieval in Digital Libraries", *ISGIS'10: Proceedings of the Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science proceedings*, p. 229–234, 2010.
- Perea-Ortega J. M., García-Cumbreras M. A., García-Vega M., Ureña-López L. A., "Comparing Several Textual Information Retrieval Systems for the Geographical Information Retrieval Task", *NLDB'08: Proceedings of the 13th international conference on Natural Language and Information Systems*, Springer-Verlag, Berlin, Heidelberg, p. 142–147, 2008.
- Peters C., "Introduction", *CLEF'01: Proceedings of the 1st Workshop Cross-Language Information Retrieval and Evaluation*, vol. 2069 of LNCS, Springer, p. 1–6, 2001.
- Pfoser D., Efentakis A., Hadzilacos T., Karagiorgou S., Vasiliou G., "Providing Universal Access to History Textbooks: A Modified GIS Case", *W2GIS'09: Proceedings of the 9th International Symposium on Web and Wireless Geographical Information Systems*, vol. 5886 of LNCS, p. 87–102, 2009.
- Sallaberry C., Baziz M., Lesbegueries J., Gaio M., "Towards an IE and IR System Dealing with Spatial Information in Digital Libraries – Evaluation Case Study", in J. Cardoso, J. Cordeiro, J. Filipe (eds), *ICEIS'07: Proceedings of the 9th International Conference on Enterprise Information Systems*, p. 190–197, 2007.
- Sallaberry C., Gaio M., Palacio D., Lesbegueries J., "Fuzzifying GIS Topological Functions for GIR Needs", *GIR'08: Proceeding of the 2nd international workshop on Geographic information retrieval*, ACM, New York, NY, USA, p. 1–8, 2008.

- Sanderson M., Kohler J., “Analyzing Geographic Queries”, *SIGIR-GIR'04: Proceedings of the Workshop on Geographic Information Retrieval at SIGIR*, 2004.
- Sanderson M., Zobel J., “Information Retrieval System Evaluation: Effort, Sensitivity, and Reliability”, *SIGIR'05: Proceedings of the 28th annual international ACM SIGIR conference*, ACM, New York, NY, USA, p. 162–169, 2005.
- Vaid S., Jones C. B., Joho H., Sanderson M., “Spatio-textual Indexing for Geographical Search on the Web”, *SSTD'05: Proceedings of the 9th international Symposium on Spatial and Temporal Databases*, vol. 3633 of *LNCS*, Springer, p. 218–235, 2005.
- Valcartier, GRID – geospatial retrieval of indexed document, Rapport technique, R&D pour la défense, Canada, 2006.
- Verhagen M., Gaizauskas R., Schilder F., Hepple M., Moszkowicz J., Pustejovsky J., “The TempEval challenge: identifying temporal relations in text”, *Lang. Resour. Eval.*, vol. 43, n° 2, p. 161–179, 2009.
- Voorhees E. M., “Evaluation by Highly Relevant Documents”, *SIGIR'01: Proceedings of the 24th annual international ACM SIGIR conference*, ACM, New York, NY, USA, p. 74–82, 2001.
- Voorhees E. M., Harman D. K., *TREC: Experiment and Evaluation in Information Retrieval*, MIT Press, Cambridge, MA, USA, 2005.
- Widlöcher A., Bilhaut F., “La plate-forme LinguaStream : un outil d’exploration linguistique sur corpus”, *TALN'05: Actes de la 12^e Conférence sur le Traitement Automatique du Langage Naturel*, 2005.
- Woodruff A. G., Plaunt C., “GIPSY: automated geographic indexing of text documents”, *J. Am. Soc. Inf. Sci.*, vol. 45, n° 9, p. 645–655, 1994.
- Yager R. R., “On ordered weighted averaging aggregation operators in multicriteria decision-making”, *IEEE Trans. Syst. Man Cybern.*, vol. 18, n° 1, p. 183–190, 1988.