LA RECHERCHE GÉOGRAPHIQUE D'INFORMATION SUR LE WEB: BESOINS ET ÉVALUATION

Bénédicte BUCHER[1], Paul CLOUGH[2], Hideo JOHO[2], Ross PURVES[3] et Awase Khirni SYED[3]

[1] Laboratoire COGIT - Institut Géographique National {benedicte.bucher@ign.fr}
[2] Department of Information Studies, University of Shef field, UK {p.d.clough;h.joho@shef field.ac.uk}
[3] Department of Geography, University of Zürich, Switzerland {rsp;sak@geo.unizh.ch}

Résumé

La recherche géographique d'information (RGI) sur le W eb est un domaine récent qui évolue rapidement. Le développement de ces applications nécessite une analyse des besoins auxquels elles doivent répondre et ainsi que des méthodes permettant d'évaluer comment elles y répondent. Cet article décrit l'analyse des besoins ayant conduit au développement de l'application SPIRIT et la méthode proposée pour son évaluation. Cette évaluation doit porter d'une part sur le compor - tement absolu du système et d'autre part sur ce comportement perçu par l'utilisateur final. Nous développons une collection particulière de documents pour faciliter la mesure des performances de SPIRIT ainsi qu'une grille de notation de la pertinence spatiale et thématique d'un document en réponse à une requête. Nous soulignons l'importance d'intégrer dans l'évaluation de systèmes RGI les interactions de l'utilisateur avec le système autant que les performances absolues du système.

Introduction

La recherche géographique d'information (RGI) sur le W eb est définie par (Larson, 1996) comme "l'aide à l'accès à des sources d'information localisée". Actuellement, l'expression *recherche d'information* (RI) renvoie généralement à la recherche de documents répondant à une requête parmi une collection importante et non structurée de documents textuels stockée sur le Web. Dans ce contexte, nous limitons la recherche géographique d'information à la recherche de documents répondant à une requête de la forme <thème, relation spatiale, localisation> parmi une collection importante et non structurée de documents textuels stockée sur le Web. Dans cette requête, le thème et la localisation sont liés par une relation spatiale d'inclusion, topologique ou directionnelle, par exemple "châteaux dans le pays de Galle", "châteaux proches du pays de Galle" ou "châteaux au nord du pays de Galle". Les travaux en RGI sont conduits par des universitaires et également par des sociétés commerciales. Par exemple, Google a développé récemment un moteur de recherche dit "local" qui s'appuie sur des annuaires commerciaux (http://local.google.co.uk/).

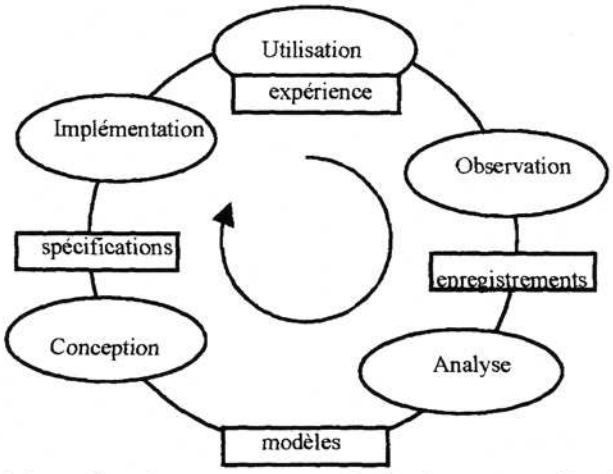

Figure 1: Activités et objets dans un processus de développement (d'après Henderson 1991, p262)

Le projet européen SPIRIT (Spatially aware information retrieval on the internet) veut développer une application de RGI qui exploite les éléments thématiques et géographiques contenus dans les documents W eb. Il adopte une démarche itérative similaire à celle illustrée sur la figure 1 qui accorde une grande importance à l'utilisateur. Ce cycle est parcouru plusieurs fois lors du développement de l'application. Avant l'implémentation, les étapes d'utilisation, observation, analyse et conception correspondent à une *analyse des besoins* conduisant à l'*établissement de spécifications fonctionnelles et de spécifications de conception*. Dès lors qu'un prototype est disponible, ces mêmes étapes correspondent à l'*évaluation*. Cet article présente l'analyse des besoins ayant présidé au développement de l'application de RGI SPIRIT et les spécifications de cette application. Nous détaillons ensuite l'évaluation de telles applications et en particulier de l'application SPIRIT.

Analyse des besoins pour SPIRIT

La méthode d'analyse des besoins utilisée dans le projet SPIRIT se compose de deux éléments. D'une part, des maquettes ont été construites et associées à des scénarios d'utilisation imaginés par les membres du projet. Ces maquettes ont été présentées à des utilisateurs potentiels qui ont ensuite exprimé leurs besoins en termes d'interactions et de fonctionnalités d'un système similaire à celui illustré par les maquettes. D'autre part, une analyse des applications existantes qui fournissent certaines de ces fonctionnalités a été conduite. Par exemple, nous avons analysé les modalités d'expression d'un lieu dans des applications comme Local Google et Multimap (www.multimap.com). Ces analyses d'une part des besoins et d'autre part des solutions partielles disponibles ont résulté en un ensemble de fonctionnalités appelant des solutions innovantes de la part du projet SPIRIT.

Un besoin clé est celui d'effectivement trouver des documents sur le Web qui soient spatialement et thématiquement pertinents en réponse à une requête composée d'un thème, d'une relation spatiale et d'une localisation. Ce besoin renvoie à plusieurs fonctionnalités. La première est l'interprétation d'un nom de lieu exprimé par l'utilisateur. Cela peut demander une levée d'ambiguïté, par exemple entre London, UK, et London, Ontario, ou encore la gestion de lieux vagues comme le sud de l'Angleterre (Purves et al., 2005). La fonctionnalité suivante consiste à transmettre la requête à un moteur de recherche qui intègre des techniques pour gérer les composantes thématiques et géographiques de la requête et ordonner les documents y répondant (V an Kreveld et al., 2004).

Les utilisateurs ont également insisté sur l'importance d'associer une carte à la présentation des réponses, surtout lorsque la pertinence spatiale d'un document n'est pas claire pour l'utilisateur. Ce besoin surgit généralement lorsque l'utilisateur ne connaît pas en détail le lieu sur lequel porte sa requête, par exemple une personne recherche des maisons d'hôte au nord de Leicester mais ne connaît pas suf fisamment la région pour savoir si une localité se trouve ef fectivement au nord de Leicester. Fournir à l'utilisateur une carte localisant sa requête ainsi que les documents W eb obtenus en réponse lui permet de juger de la *pertinence spatiale* des résultats, c'est-à-dire d'évaluer visuellement la relation spatiale spécifiée dans sa requête entre le lieu exprimé et les *empreintes géographiques* des documents. Les utilisateurs ont aussi exprimé le besoin de préciser leur requête de façon itérative en spécifiant sur la carte une nouvelle zone d'intérêt.

Certains utilisateurs recherchent enfin non seulement des documents Web mais aussi des données géographiques. Par exemple, un utilisateur recherchant des informations sur "la randonnée dans les Alpes" aimerait obtenir des données altimétriques sur cette région.

L'application SPIRIT

L'analyse des besoins menée dans SPIRIT a conduit à la définition de spécifications initiales pour notre application. L'application se compose de plusieurs modules. L 'interface utilisateur supporte l'expression de la requête et la présentation des résultats. Le moteur de recherche et le module de classement recherchent et classent les documents répondant à une requête de la forme <thème, relation spatiale, localisation>. Enfin, une ontologie géographique stocke et gère les connaissances concernant la sémantique et la géométrie des lieux (Jones et al., 2004). L 'utilisateur interagit avec cette architecture uniquement via l'interface utilisateur (Purves et al., 2005).

L'interface permet à l'utilisateur d'exprimer un besoin de deux façons : soit en formulant une requête structurée soit en dessinant une requête graphique. Dans le cas d'une requête structurée, illustrée sur la figure 2, l'utilisateur spécifie un nom de lieu, comme Edinburgh. Dans le cas d'une requête graphique, il dessine un polygone sur la carte proposée par l'interface. Les autres modules composant SPIRIT, comme le module de classement, implémentent également plusieurs méthodes. Par exemple, le module de classement implémente plusieurs mesures de pertinence et peut classer les résultats soit uniquement selon leur pertinence thématique soit selon diverses combinaisons de la pertinence spatiale et de la pertinence thématique.

Lors de l'évaluation, il sera nécessaire non seulement d'évaluer la qualité des résultats obtenus avec ces différentes modalités ou méthodes, mais aussi de les comparer pour déterminer quelles techniques sont les plus ef ficaces.

Figure 2: Interface de SPIRIT supportant l'expression de la requête structurée <castles, in , wales>. SPIRIT propose en réponse une liste ordonnée de documents associée à une carte les localisant.

L'évaluation d'applications de recherche géographique d'information

La recherche géographique d'information (RGI) est un domaine en constante évolution. Peu de systèmes de RGI sont basés sur les techniques de recherche d'information (RI) car ils sont pour la plupart des systèmes de « recherche géographique d'information géographique » et non de recherche de documents Web. A notre connaissance, il n'y a pas de méthode d'évaluation proposée pour des applications comme SPIRIT qui soit adaptée à ces deux aspects : la recherche d'information parmi une collection non structurée de documents textuels et le traitement spécifique de l'information géographique lors de cette recherche. Cela dit, comme de tels systèmes émergent peu à peu, il sera de plus en plus crucial de disposer de méthodes d'évaluation suffisamment génériques pour pouvoir s'appliquer à tous, permettre de les comparer et de déterminer les techniques les plus efficaces.

Par contre, l'évaluation a été beaucoup étudiée dans le domaine de la recherche d'information (RI). La complexité de l'évaluation dans ce domaine réside essentiellement dans la difficulté d'appréhender et de mesurer la pertinence d'un document pour un utilisateur formulant une requête. En effet, cette pertinence est souvent subjective, voire controversée (Saracevic, 1975). Deux stratégies d'évaluation ont été peu à peu développées dans la littérature en RI (Spark Jones and Willett, 1997:167-174): l'évaluation orientée système et l'évaluation orientée utilisateurs.

L'évaluation orientée système vise à mesurer les performances du système de façon la plus standard et objective possible (Borlund, 2003, Van Rijsbergen, 1979). Cette évaluation permet de comparer des applications RI ou différentes implémentations de modules d'une même application. Elle met en place un banc d'essai simulant des tâches de recherche d'information en l'absence d'utilisateurs. Ce banc d'essai consiste en une ressource standard appelée collection test, qui se compose des éléments suivants :

- un ensemble de documents Web représentant un domaine ou le Web, D,
- un ensemble de requêtes représentant les tâches utilisateurs de façon réaliste et contrôlable (Peters, 2001: 1069), R,
- des notes de pertinence pour chaque document Web et chaque requête, P. La construction de ces notes s'appuie sur une grille de notation qui doit guider la définition de notes de pertinence de façon la plus objective et générique possible. L'attribution de ces notes est l'étape la plus longue de la constitution d'une collection test.

Ainsi faite, une collection test simule le résultat idéal de tâches utilisateurs. Ses deux premiers éléments, D et R, peuvent être utilisés pour reproduire ses tâches à l'aide d'une application de RI dont on veut mesurer les performances.

Les mesures sont obtenues en comparant les documents trouvés par l'application au sein de D pour répondre à chaque requête de R avec les notes de pertinences appartenant à P.

Deux grandes mesures résument les performances d'une application de RI et servent de base au calcul d'autres mesures : la précision et le rappel. Ces mesures sont classiquement définies pour une collection dans laquelle les notes de pertinence sont binaires : pour une requête et un document donnés, le document est pertinent ou ne l'est pas. Dans ce contexte, la précision mesure le nombre de documents pertinents retournés par l'application parmi les documents retournés par cette application. A ce stade, il faut souligner qu'une application peut retourner un nombre très important de documents et que les documents les moins bien classés seront rarement consultés par l'utilisateur . Aussi, une mesure plus intéressante est la "précision à n", où n vaut par exemple 10. La précision à n vaudra donc : le nombre de documents pertinents parmi les 10 premiers documents trouvés par l'application. Le rappel mesure lui le nombre de documents retournés parmi les documents pertinents. Des mesures de précision et rappel ont été proposées pour s'adapter à des notes de pertinence plus détaillées (Kekäläinen and Järvelin, 2002). Cela permet une plus grande expressivité, en particulier dans le domaine de la recherche d'information parmi une collection de documents XML structurés (Kazai et al., 2004), ou dans le domaine de la recherche d'images (Clough et al., 2005).

La conception de collections test standard a commencé il y a 40 ans sous l'impulsion de Cleverdon (1967) et est depuis lors un élément de référence de l'évaluation d'applications de RI, mis en oeuvre par exemple dans les campagnes de la Text REtrieval Conference (TREC). Dans de telles campagnes d'évaluation, les participants utilisent la même collection test –ensemble de documents W eb et ensemble de requêtes- pour mesurer les performances de leurs systèmes de façon comparable. Ces campagnes étudient également les interactions utilisateurs.

L'évaluation orientée utilisateurs impliquent les utilisateurs, au contraire de l'évaluation orientée système. Elle prend son sens lorsque l'application a une interface utilisateurs. Ces dernières années, le développement d'applications de RI offrant une plus grande interactivité a remis en cause la stratégie purement orientée système pour évaluer ces applications (Borlund, 2003). Il est maintenant conseillé d'adopter une stratégie complémentaire orientée utilisateurs pour mesurer en quoi l'application, globalement, favorise le processus de recherche d'information mené par l'utilisateur. Cette évaluation étudie d'une part des éléments spécifiques de l'interface et d'autre part l'utilisabilité globale du système par les utilisateurs, c'est-à-dire la possibilité pour les utilisateurs de réaliser leurs tâches de recherche d'information et la facilité avec laquelle ils la réalisent. Borlund (2003) propose d'évaluer des systèmes de RI interactifs en s'appuyant sur un scénario d'utilisation qui couvre les principales tâches utilisateurs. L'utilisateur suit le scénario et évalue la pertinence des documents qu'il obtient. Cette pertinence est relative au contexte défini par le scénario. Il est également possible de mettre en place des interviews pour approfondir par exemple l'évaluation de l'utilisabilité du système. Une littérature importante existe dans ce domaine, que ce soit sur l'évaluation de logiciels en général ou de systèmes de RI (e.g. Bawden, 1990). En définitive, concernant l'évaluation orientée utilisateurs, il n'existe pas de forte spécificité liée au domaine de la RI et cette évaluation se fait comme pour les logiciels en général.

L'évaluation dans SPIRIT

L'évaluation dans SPIRIT se compose de l'évaluation orientée système et de l'évaluation orientée utilisateurs. L 'évaluation orientée utilisateur s'appuie sur des scénarios comme proposé par (Borlund 2003) ainsi que sur un questionnaire. Les scénarios et le questionnaire donnent une place importante aux activités de l'utilisateur relatives à l'information spatiale, comme l'expression des aspects spatiaux de sa requête, l'appréhension des aspects spatiaux d'un document Web ou l'interprétation du classement spatial.

L'évaluation orientée système présente plus de spécificités. Pour la mettre en place, il faut définir une collection test c'est-à-dire : un ensemble de documents W eb, un ensemble de requêtes, une grille de définition de notes de pertinence d'un document relativement à une requête et les notes de pertinence affectées à chaque document pour chaque requête. La conception de cette collection test doit prendre en compte plusieurs facteurs :

- l'évaluation doit mesurer tout particulièrement la gestion des aspects spatiaux de la recherche d'information,
- l'évaluation doit être adaptée aux limites du prototype SPIRIT en termes de documents W eb et de données géographiques. En effet, les connaissances géographiques de SPIRIT sont organisées dans l'ontologie géographique construite à partir de données géographiques. L 'alimentation de cette ontologie a été limitée aux données géographiques que nous avons pu acquérir dans le contexte du projet. Par ailleurs SPIRIT ne fonctionne pas sur le W eb mais sur une copie partielle du W eb acquise en début de projet auprès de Google. Nous détaillons ci-dessous la définition des éléments composant la collection test de SPIRIT .

Les documents W eb retenus dans la collection test, c'est-à-dire l'élément D introduit dans la définition d'une collection test, doivent d'une part être représentatifs des documents sur lesquels SPIRIT fonctionnera et d'autre part être en nombre suffisamment restreint pour que la production des notes de pertinence soit possible. Ces documents ont été extraits d'une copie partielle du W eb d'1 Terabyte acquise auprès de Google au début du projet SPIRIT (Joho and Sanderson, 2004). En prévision de l'évaluation des capacités spatiales de SPIRIT, nous avons pris soin de garder dans cet ensemble de documents W eb les réponses à des requêtes impliquant une relation spatiale. Par exemple, concernant la requête « pubs near Glencoe », nous avons intégré dans cette collection les documents trouvés par un moteur de recherche classique fonctionnant sur cette copie pour des requêtes textuelles « pub, x » où x sont les noms de localités proches de Glencoe. Les requêtes utilisées pour construire cette collection test sont bien plus nombreuses que celles utilisées pour l'évaluation. Pour les générer , nous avons dans un premier temps utilisé les noms des 200 plus grandes villes du Royaume Uni et gardé pour chacune les 50 premiers documents renvoyés par le moteur de recherche textuel. Le résultat consistait en 9 010 documents et 85 MB de texte. Dans un deuxième temps, nous avons utilisé des requêtes correspondant à des besoins d'utilisateurs tels "Arts festivals dans Edinburgh" qui nous semblaient repré-

sentatifs des besoins pour un système comme SPIRIT . Nous avons choisi de privilégier des localisations dans le Royaume Uni tout en utilisant d'autres localisations comme Montreux en Suisse.

La définition des requêtes composant la collection test de SPIRIT, c'est-à-dire l'élément R introduit dans la définition d'une collection test, a été faite en considérant les requêtes pour lesquelles les moteurs classiques n'apportent pas une satisfaction suffisante à l'utilisateur, c'est-à-dire les requêtes correspondant à un besoin de moteur de recherche géographique d'information. Il s'agit de requêtes de la forme <thème, relation spatiale, localisation> ayant les propriétés suivantes :

- Le nom de lieu est aussi un mot commun davantage utilisé, comme Battle (nom de lieu anglais signifiant par ailleurs bataille).
- Le nom de lieu est ambigu, comme Paris au Texas.
- Le nom de lieu utilisé dans la requête est peu susceptible d'être utilisé dans les documents réponse, comme dans <blog, près de, Vélieux>.
- Le nom de lieu renvoie à une région imprécise, comme le Sud de la France. Dans ce cas, l'application doit avoir une intelligence spatiale complexe pour formaliser l'empreinte géographique de la requête.
- Le thème de la requête est aussi un nom de lieu, comme Forêt noire.
- La relation spatiale n'est pas une relation d'inclusion, comme <maison d'hôte, au nord de, Auxerre>.

Une première grille de définition de notes de pertinence d'un document relativement à une requête a été proposée. Elle est présentée ci-dessous. Une note est composée de deux variables : la pertinence thématique et la pertinence spatiale. Chaque variable est ternaire pour distinguer plusieurs niveaux de pertinence. Notons que la pertinence spatiale est dépendante non seulement de la localisation mais aussi du thème car l'échelle d'une carte « locale » sur laquelle situer le document est dépendante de la granularité du thème. Ainsi selon que je recherche des informations de randonnée ou de météorologie, le niveau de détail des informations de localisation dont l'utilisateur a besoin ne sera pas le même.

Pertinence thématique

1. Le document contient de l'information pertinente relativement à la requête et permet à l'utilisateur d'appréhender cette pertinence sans l'aide d'autres connaissances. Autrement dit, le lien du document avec le thème est suffisamment clair et ne nécessite pas d'expertise pour être établi.
2. Le document contient de l'information pertinente mais il faut consulter d'autres ressources pour former un jugement sur cette pertinence. Par exemple, le document mentionne le thème mais se contente de pointer vers une ressource sur ce thème. Ou encore le document a un lien avec le thème mais ce lien n'est pas simple à appréhender pour un non expert et il faut s'aider de ressources plus simples pour le comprendre.
3. Le document n'a pas de lien avec le thème de la requête.

Pertinence spatiale

1. Le document fait référence à une localisation répondant aux critères de la requête et cette localisation est suffisamment détaillée dans le document pour que l'utilisateur puisse la retrouver sur une carte locale.
2. Le document fait référence à une localisation répondant aux critères de la requête mais il n'y a pas suffisamment d'information de localisation pour que l'utilisateur puisse la retrouver sur une carte locale.
3. Le document ne fait pas référence à une localisation répondant aux critères de la requête.

Tableau 1: Grille initialement définie dans SPIRIT pour noter la pertinence d'un document W eb en réponse à une requête de la forme <thème, relation spatiale, localisation>.

Nous avons testé l'utilisabilité de cette grille en nous appuyant sur 5 requêtes décrites dans le tableau 2. Pour chacune, les 10 premiers documents renvoyés par SPIRIT ont été proposés à 11 évaluateurs qui ont dû les noter conformément à la grille. Précédemment, des exemples de jugements ont été donnés aux évaluateurs pour illustrer la grille.

1.	Caving in Derbyshire (UK)
2.	Castles in Wales (UK)
3.	Skiing near Glencoe (UK)
4.	Art festivals in Edinburgh (UK)
5.	Music in Montreux (Switzerland)

Tableau 2: Requêtes utilisées pour tester la grille de notation de pertinence.

Ces évaluateurs ont répondu à un questionnaire pour indiquer leur compréhension de la grille et commenter l'utilisation qu'ils en ont faite. En général, ils ont trouvé la grille simple à comprendre mais ont eu des dif ficultés à attribuer des notes de pertinence spatiale. Il est ainsi apparu logiquement qu'un évaluateur doit avoir une bonne connaissance de la localisation de la requête pour attribuer une note de pertinence spatiale lorsque cette note vaut 2. Par exemple, il existe une piste de ski près de Glencoe appelée « White Corries ». Si la localisation Glencoe ne figure pas sur un document relatif à cette piste, il faudra lui attribuer une note de pertinence spatiale de valeur 2 mais seul un évaluateur connaissant le voisinage de Glencoe et en particulier cette piste de ski le saura. Ainsi, pour construire les notes de pertinence il faut que les évaluateurs aient une bonne connaissance de la localité sur laquelle ils ef fectuent leur requête. Il faut également tenir compte des indices de localisation disponibles dans le document, qui ne sont pas toujours sous la forme d'un nom de lieu ou d'une adresse. Par ailleurs, les évaluateurs ont trouvé inutile la distinction entre une pertinence thématique de 2 et de 1.

Nous avons par ailleurs mesuré la cohérence entre les notes attribuées par les 11 évaluateurs en utilisant une mesure de Cohen Kappa. La valeur de cohérence obtenue pour les notes spatiales et pour les notes thématiques est de 95%, ce qui est largement satisfaisant.

Conclusion et perspectives

Cet article a décrit l'analyse des besoins et une démarche de mise en place d'une méthode d'évaluation pour une application de recherche géographique d'information, SPIRIT. L'évaluation de telles applications nécessite l'adaptation de techniques existantes d'évaluation d'application en RI. Cela comprend en particulier le développement de nouvelles grilles de notation de pertinence. La définition de ressources standard dans ce domaine aiderait à déterminer quelles sont les techniques les plus ef ficaces et donc faciliterait le progrès en RGI. Dans le contexte de SPIRIT, une première grille de notation de pertinence a été proposée et testée. Nous avons utilisé pour l'évaluation finale une nouvelle grille ne conservant que deux niveaux de pertinence thématique (pertinent / non pertinent) et conservant les 3 niveaux de pertinence spatiale de la grille initiale. Des mesures de précision et rappel de SPIRIT ont pu être faites par la suite à l'aide de cette collection test et ont montré l'apport des techniques de SPIRIT à la RGI par rapport à un moteur de recherche purement textuel. De plus, ces applications impliquent généralement une forte interactivité avec l'utilisateur de sorte que leur évaluation ne peut faire l'économie d'une évaluation orientée utilisateurs. Celle-ci peut être menée en s'appuyant sur des scénarios d'utilisation et un questionnaire. A l'avenir, nous pensons que la formalisation de la notion de pertinence d'une information en RGI devra s'appuyer sur la modélisation d'activités spatiales et de l' « affordance » des objets géographiques dans ces activités (Jordan et al., 1998).

Remerciements

Ces travaux sont financés en partie par le projet européen No. IST -2001-35047 (SPIRIT) et par le BBW Suisse (01.0501).

Références

Bawden, D. 1990. User-orientated evaluation of information systems and services», Gower Publishing company .

Borlund, P. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. In: *Information Research*, vol. 8, no. 3, paper no. 152.

Cleverdon, C. The Cranfield test on index language devices: In Sparck Jones, K. & W illett, P. eds. (1997) *Readings in Information Retrieval*. San Francisco: Morgan Kaufmann. Pp. 47-59.

Clough, P., Mueller, H. and Sanderson, M. (2005), The CLEF 2004 Cross Language Image Retrieval Track, *In Multilingual Information Access for Text, Speech and Images: Results of the Fifth CLEF Evaluation Campaign,* Eds (Peters, C., Clough, P., Gonzalo, J., Jones, G., Kluck, M. and Magnini, B.), Lecture Notes in Computer Science (LNCS), Springer , Heidelberg, Germany (in print).

Henderson A. 1991. A Development Perspective on Interface Design and Theory in J.M. Carroll (Ed): *Designing Interaction: Psychology at the Human Computer Interface* . Cambridge, Cambridge University Press: 254-268.

Joho, H., and Sanderson, M. 2004. The SPIRIT Collection: an overview of a large web collection. *SIGIR Forum*, 38(2).

Jones, C.B., Abdelmoty, A.I., Finch, D., Fu, G. & Vaid, S. 2004. The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. In *Proceedings of the 3rd International Conference on Geographic Information Science* (GIScience 2004), Maryland, USA. LNCS.

Jordan, T. M. Raubal, B. Gartrell, and M. Egenhofer. 1998. An Affordance-Based Model of Place in GIS. in: T. Poiker and N. Chrisman (Eds.), 8th Int. Symposium on Spatial Data Handling, SDH'98, V ancouver, Canada, pp. 98-109.

Kazai, G., Lalmas, M., De Vries, P. 2004. The overlap problem in content-oriented XML retrieval evaluation. SIGIR 2004: 72-79

Kekäläinen, J. & Järvelin, K. 2002. Using graded relevance assessments in IR evaluation. *Journal of the American Society for Information Science and Technology 53*(13), 1120-1129.

Larson, R.R. 1996. Geographic Information Retrieval and Spatial Browsing. In *GIS and Libraries: Patrons, Maps and Spatial Information*, Linda Smith and Myke Gluck, Eds., University of Illinois.

Purves, R.S., Clough, P. and Joho, H. 2005. Identifying imprecise regions for geographic information retrieval using the web. In *Proceedings of GISRUK*, 2005.

Saracevic, T. 1975. Relevance: a review of and a framework for the thinking on the topic. *Journal of the American Society for Information Science*, vol. 26: 321-343

Sparck Jones, K. & Willett, P. eds. 1997. *Readings in Information Retrieval*. San Francisco: Morgan Kaufmann.

Van Kreveld, M., Reinbacher, I., Arampatzis, A. and Van Zwol, R. 2005. Multi-Dimensional Scattered Ranking Methods for Geographic Information Retrieval. *Geoinformatica*, 9,1, 61-84.

Van Rijsbergen, C. J. Information Retrieval, Butterworth-Heinemann, Newton, MA, 1979 (http://www .dcs.gla.ac.uk/Keith/ Preface.html)

Voorhees, E.M. 2001. Overview of TREC 2001. In Proceedings of TREC 2001, NIST.