

## **Nouveautés de la technologie SIG**

Bulletin d'information de l'OSIG 1-2013

### **La proportion des informations à référence spatiale est-elle vraiment de 80% ?**

#### **Introduction**

La thèse selon laquelle 80% de toutes les informations ont une référence spatiale est largement accréditée dans le milieu des SIG. Apparue dans les années 1980 en même temps que les premiers systèmes d'information géographique, elle est brandie depuis lors comme un argument marketing par une frange importante des acteurs du secteur. Pourtant, si l'on se met en quête des origines de cette thèse, il est impossible de trouver trace de la moindre étude confirmant ou infirmant empiriquement la proportion ainsi avancée. Les sources les plus anciennes semblent indiquer qu'elle se rapportait initialement aux données communales, parce que ces dernières comportent de multiples références à des données telles que des adresses, des coordonnées ou des numéros permettant une identification spatiale (des numéros de parcelles ou des codes postaux par exemple). Au fil du temps, cette thèse a cependant été généralisée (à l'excès) et a été étendue à tous types de contextes. Les auteurs de publications scientifiques n'ont d'ailleurs pas manqué de s'en servir. Une étude empirique actuellement menée par la chaire de communication cartographique de l'université technique de Dresde vise donc à combler la lacune ainsi pointée.

#### **Wikipedia, base de recherche et réseau**

Pour que l'étude puisse présenter un caractère aussi général que possible, il a fallu trouver une base regroupant une grande quantité d'informations sans pour autant sous-représenter ou surreprésenter l'une ou l'autre discipline ou spécialité. Notre choix s'est donc porté sur l'encyclopédie en ligne Wikipedia. On peut la modéliser comme un réseau. Présentée sous forme de graphe orienté, les liens entre les différents articles en constitueraient les arcs et les articles eux-mêmes formeraient les sommets. Actuellement, près de 18% des articles de la version en allemand de Wikipedia comportent une identification explicite par des coordonnées géographiques. C'est par exemple le cas de ceux consacrés aux villes, aux montagnes ou aux fleuves. On peut donc dire d'eux qu'il s'agit d'«articles géospatiaux». Notre approche s'apparente à celle des réseaux sociaux qui distinguent des contacts de premier, deuxième, ... niveau, puisque nous définissons des articles à référence spatiale de premier, deuxième, ... niveau. Cette classification s'est fondée sur la détermination, pour chacun des articles du réseau, du chemin le plus court vers un «article géospatial».

#### **Validation cognitive**

L'intégration de coordonnées explicites et de liens permettant de naviguer d'un article à l'autre témoignent d'un processus cognitif, si bien qu'il est légitime d'estimer que le niveau d'un article au sein du réseau, établi tel qu'indiqué par voie d'analyse, résulte lui aussi d'un processus cognitif. Une étude empirique a par ailleurs cherché à déterminer le degré de corrélation existant entre le niveau de référence spatiale ainsi défini et l'évaluation qu'en faisait un échantillon de personnes interrogées à ce sujet. Il s'agissait surtout de répondre à la question suivante: à quel niveau la séparation entre informations à référence spatiale / sans référence spatiale est-elle réalisée par la perception humaine? Un échantillon de 2200 articles de Wikipedia a servi à cette fin. Il a été demandé aux personnes participant à cette étude de classer chacun de ces articles dans l'une des catégories proposées et fréquemment utilisées (référence spatiale directe, référence spatiale indirecte, aucune référence spatiale) sur la seule base de son titre.

#### **Résultats**

Du fait de la forte interconnexion des articles de Wikipedia, l'approche réseau n'a dégagé que quatre niveaux principaux de référence spatiale: le niveau 0 concerne 17,6% des articles, le niveau 1 en touche 60,9%, le niveau 2 en regroupe 20,7% et le niveau 3 n'en rassemble que 0,8%. Les trois niveaux de référence les plus pertinents – soit les niveaux 0, 1 et 2 – ont fait

l'objet d'un examen plus attentif. Au niveau 0, on trouve des articles tels que ceux consacrés à Zurich, au Jura souabe ou à Teldafax, une holding très connue en Allemagne. Au niveau 1, on citera, à titre d'exemples, les articles portant sur Friedrich Schiller, la Révolution française ou l'aluminium. Au niveau 2 enfin, les articles concernent des notions plus abstraites telles que les fonctions linéaires, la musique pop synthétique ou encore la poudre à canon. L'approche réseau associe donc un niveau de référence spatiale (plus ou moins élevé) à chacun des articles de Wikipedia, de sorte que 100% des informations sont d'une certaine manière à référence spatiale. Les questions posées aux personnes sondées ont donc servi à déterminer la proportion des informations auxquelles un individu associe spontanément une référence spatiale. Après dépouillement, il s'est avéré que 90% des articles de niveau 0 se sont vus attribuer une «référence spatiale directe», que les articles de niveau 1 se partageaient dans des proportions sensiblement égales entre les catégories «référence spatiale indirecte» (39%) et «aucune référence spatiale» (44%) et que les articles de niveau 2 étaient majoritairement rangés dans la catégorie «Aucune référence spatiale» (72%). En combinant les deux approches, on peut donc estimer que 57% des articles de la version en allemand de Wikipedia ont une référence spatiale (directe pour 27% d'entre eux, indirecte dans 30% des cas) et que 43% des articles n'ont «aucune référence spatiale». En arrondissant, on peut donc en conclure que la thèse dite des 80% devrait être débaptisée et renommée thèse des 60%.

### **Discussion**

Certaines restrictions sont attachées à la méthode employée. On peut par exemple mentionner le fait que les articles de Wikipedia sont souvent rédigés à un niveau conceptuel, si bien que la notion de «maison» est certainement perçue de manière bien différente, en termes de référence spatiale, que l'instance concrète «Maison Blanche (Washington D.C.)». En outre, Wikipedia est une encyclopédie, ce qui la distingue fondamentalement de la source dont est issue la thèse des 80%, à savoir les données communales. Il convient donc de ne pas donner une portée trop générale aux résultats de la présente étude.

### **Publications**

Le lecteur est renvoyé aux publications suivantes pour une présentation complète des résultats:

- Hahmann, S.; Burghardt, D. (2012): Forschungsergebnisse zur Frage: Haben 80% aller Informationen einen Raumbezug? Dans: gis.SCIENCE - Die Zeitschrift für Geoinformatik, n°3, 25, Wichmann, p.101–108
- Hahmann, S.; Burghardt, D.: How much information is geospatially referenced? Networks and Cognition. Dans: International Journal of Geographical Information Science, en cours d'impression, DOI: 10.1080/13658816.2012.743664
- Des pré-impressions de ces deux publications sont disponibles sous:
  - [http://kartographie.geo.tu-dresden.de/geospatial\\_experiment/Hahmann\\_Burghardt\\_gisscience\\_Preprint.pdf](http://kartographie.geo.tu-dresden.de/geospatial_experiment/Hahmann_Burghardt_gisscience_Preprint.pdf)
  - [http://kartographie.geo.tu-dresden.de/geospatial\\_experiment/Hahmann\\_Burghardt\\_IJGIS\\_Preprint.pdf](http://kartographie.geo.tu-dresden.de/geospatial_experiment/Hahmann_Burghardt_IJGIS_Preprint.pdf)

Groupe de travail technologie SIG de l'OSIG

[technologie@sogi.ch](mailto:technologie@sogi.ch)

Stefan Hahmann (collaborateur scientifique et doctorant)

Dirk Burghardt (professeur de communication cartographique)

Les deux auteurs travaillent actuellement à l'Institut de cartographie de l'université technique de Dresde.