

Résumé du stage

Étant encore incertaine quant à ma future orientation professionnelle, j'ai décidé de profiter du stage destiné à valider ma troisième année de licence afin de découvrir le monde de la recherche. Pour cela j'ai contacté Guillaume Cabanac qui m'a proposé le stage en scientométrie – étude quantitative de la science par une démarche scientifique – que je vais vous décrire ci-dessous.

La première partie de ce stage – destinée à me familiariser avec le vocabulaire de la recherche scientifique et à me permettre d'acquérir quelques méthodes scientométriques – consistait à valoriser Inforsid, un congrès réunissant chaque année depuis 1983 les chercheurs en Sciences de l'Information à l'aide d'une base de données contenant, pour chaque édition du congrès, son comité de programme – les chercheurs chargés de sélectionner les présentations –, la ville l'ayant accueilli et les articles présentés.

Après intégration des données des éditions 2012 et 2013, j'ai tout d'abord réalisé deux cartes, une de France et une d'Europe, mettant en évidence les pôles principaux du congrès, qui s'avéraient être Paris et Toulouse. Afin de visualiser les principaux thèmes du congrès et leur évolution dans le temps j'ai ensuite créé trois nuages de mots (un nuage par décennie) à partir des concepts les plus présents dans les titres des articles présentés, mots ou expressions de plusieurs mots – telles que « base de données » ou « recherche d'information ». J'ai ensuite réalisé une carte représentant 77 revues scientifiques du domaine sciences de l'information sur laquelle étaient mis en évidence celles comprenant des membres d'Inforsid dans leur comité de rédaction – les chercheurs chargés de sélectionner les articles à publier dans le journal. Enfin j'ai réalisé une carte représentant les différentes villes ayant successivement accueilli le congrès.

La seconde mission du stage était une étude de genre des membres de comités de rédaction des 77 revues scientifiques évoquées précédemment – c'est à dire une analyse des différences et des similitudes entre les membres masculins et féminins – destinée à alimenter la rédaction d'un article scientifique. Je disposais des bibliographies de 2850 membres – 422 femmes, 2402 hommes et 26 membres dont le sexe n'avait pas pu être déterminé. J'ai calculé plusieurs indicateurs, telles que la production – mesurant la quantité de documents produits par le membre –, l'homophilie – mesurant la propension du membre à collaborer avec des femmes – ou le ϕ -index – mesurant la capacité du membre à maintenir des collaborations.

Après avoir réalisé que notre calcul de l'homophilie était incorrect à cause d'un manque d'information concernant le genre des coauteurs – et incalculable à cause de l'énorme quantité de données que nous aurions à annoter manuellement – j'ai décidé de me documenter en consultant les articles scientifiques existants traitant d'étude de genre. Ceux-ci m'ont décidé à orienter mon étude vers une comparaison de générations : j'ai divisé mon échantillon en deux – les membres ayant publié leur premier article avant 2000 et ceux l'ayant publié en 2000 ou après – et observé l'évolution des différences entre hommes et femmes. J'ai pour cela utilisé la productivité et le ϕ -index, et dans les deux cas on constate que les différences observées sur l'ancienne génération – les femmes produisant moins que les hommes et ayant plus de mal à collaborer ou à maintenir des collaborations avec leurs pairs – tendent à disparaître dans la nouvelle génération – pour laquelle, sur ces métriques, aucune différence significative n'est visible entre hommes et femmes.

Au moment où je rédige ce résumé, cette étude de genre n'est pas encore finalisée mais je pense orienter mon article vers cette différence entre générations.

Ce stage aura été une expérience profondément enrichissante pour moi. J'ai eu la chance de réaliser un travail de recherche avec un maître de stage ouvert à mes suggestions et toujours présent pour me conseiller. De plus j'ai pu découvrir de nouveaux outils, concepts et méthodes qui pourront m'être utiles dans le futur.

Internship summary

Being uncertain about my future career, I decided to experience the world of research during the internship *finishing my undergraduate degree*. For this purpose I contacted Guillaume Cabanac who offered me the *internship in scientometry* – quantitative study of science by a scientific approach – which I will describe below.

The first part of the placement – designed to familiarize myself with the language of scientific research and allow me to gain some scientometric methods – was to promote Inforsid, a congress gathering researchers in Information Science since 1983, using a database containing, for each edition of the congress, the program committee – the researchers responsible for selecting presentations – the city having hosted and papers presented.

After *having integrated data from 2012 and 2013 editions*, I first performed two maps, one from France and one from Europe, highlighting *the congress's main contributors*, which turned out to be Paris and Toulouse. To view the main themes and their evolution in time I then created three word clouds (one cloud per decade) from the most recurrent concepts in the titles of presented papers, words or term of several words – such as "information retrieval". I later produced a chart representing 77 scientific journals from the field Science of information upon which were highlighted those comprising members Inforsid in their editorial board - the researchers responsible for selecting articles for publication in the newspaper. Finally I made a map showing the cities having successively hosted the conference.

The second mission of the traineeship was a gender study – ie an analysis of the differences and similarities between male and female members – using as data the members of the editorial boards of the 77 scientific journals mentioned above intended to supply the drafting of a scientific article. I had 2850 *members' bibliographies* – 422 women, 2402 men and 26 members whose gender could not be determined. I calculated several indicators, such as production - measuring the amount of material produced by the member –, homophily – measuring the propensity of members to collaborate with women – or ϕ -index – measuring the ability of the member to maintain collaborations.

After realizing our calculation of homophily was wrong because of a lack of information about the gender of coauthors – and incalculable because of the huge amount of data that we would have to manually annotate – I decided to document myself by reading existing scientific papers dealing with gender study. These ones gave me the idea to direct my study to a comparison of generations : I divided my sample to create two sub-groups – members who published their first article before 2000 and those having published it in 2000 or later – and studied the evolution of differences between men and women. I used productivity and ϕ -index for this study, and in both cases we find that the differences existing in the older generation – women producing less than men and having more difficulty working or maintaining collaborations with their peers – tend to disappear in the next generation – on which, on these metrics, no significant difference was noted between men and women.

At the time I write this summary, this gender study is not yet finalized but I think I'll focus my article on this difference between generations.

This course has been a deeply rewarding experience for me. I had the chance to do research work with a tutor open to my suggestions and always there to advise me. Furthermore I discovered new tools, concepts and methods that will be useful to me in the future.