

# A Disciplinary Analysis of Internet Science

Clare Hooper<sup>1</sup>, Bruna Neves<sup>1</sup>, and Georgeta Bordea<sup>2</sup>

<sup>1</sup> The University of Southampton IT Innovation Centre, Southampton, UK  
[cjh@it-innovation.soton.ac.uk](mailto:cjh@it-innovation.soton.ac.uk)

<sup>2</sup> Insight, National University of Ireland, Galway  
[bordea.georgeta@deri.org](mailto:bordea.georgeta@deri.org)

**Abstract.** Internet Science is an interdisciplinary field. Motivated by the unforeseen scale and impact of the Internet, it addresses Internet-related research questions in a holistic manner, incorporating epistemologies from a broad set of disciplines. Nonetheless, there is little empirical evidence of the levels of disciplinary representation within this field.

This paper describes an analysis of the presence of different disciplines in Internet Science based on techniques from Natural Language Processing and network analysis. Key terms from Internet Science are identified, as are nine application contexts. The results are compared with a disciplinary analysis of Web Science, showing a surprisingly low amount of overlap between these two related fields. A practical use of the results within Internet Science is described. Finally, next steps are presented that will consolidate the analysis regarding representation of less technologically-oriented disciplines within Internet Science.

**Keywords.** Internet Science; disciplinary analysis; interdisciplinarity; bibliometrics; natural language processing

## 1 Introduction

Internet Science involves interdisciplinary collaboration to deepen our understanding of the Internet as a societal and technological artefact, whose evolution is increasingly intertwined with that of human societies [1]. Like the fields of Web Science and Human-Computer Interaction [2], Internet Science's interdisciplinarity is a key strength but at times a challenge: conducting successful collaborations across disciplinary boundaries can be difficult, demonstrated by efforts to discuss [3] and address [4] [5] such issues.

This paper builds on previous work in which we established a method to empirically evaluate the presence of disciplines in an interdisciplinary field based on a Natural Language Processing (NLP) analysis of a corpus of data [6]. This method has been applied in the related field of Web Science, but here we focus on its application to Internet Science, identifying key topics and application contexts of the field.

The benefits of such an analysis are wide-ranging, relating to:

1. **Quality:** work by Lungeanu [7] has shown that collaborator diversity is positively correlated with the quality of work.
2. **Credibility:** the Internet Science community claims to be interdisciplinary [1], but at this time has little evidence to back this up.
3. **Insight:** By identifying over- and under-representation of disciplines in Internet Science, we become able to take corrective action if needed. For example, we can reach out to under-represented communities by writing targeted calls for papers and co-locating Internet Science workshops and conferences with appropriate events.
4. **Communication:** By better understanding what disciplines are active within Internet Science and what we mean by the term ‘Internet Science’, we can communicate better as a community. This has positive benefits for internal and external communication, as well as for curriculum planning.

The analysis in this paper is made possible by the EU Network of Excellence in Internet Science (EINS), a European-funded research instrument to coordinate and integrate European research in the area of Internet Science [1]. Among other things, EINS has fostered the formation of a group of academics actively working in this area, and this group have published widely on a variety of Internet Science topics. These publications are the key input to the analysis described in this paper.

This paper is structured as follows: We first introduce the area of bibliometrics and the use of NLP to analyse a corpus of data, and describe our previous work in this area. We then describe the disciplinary analysis method, from initial data gathering, through processing and visualising the data, to conducting a ‘community analysis’ of that data. We then present our results, including an analysis of application contexts within the resultant graph. After discussing these results, including a comparison of Internet Science and Web Science results as well as a discussion of how the results have been applied within current Internet Science work, we discuss avenues for future work and present our conclusions.

## 2 Background and Related Work

### 2.1 Bibliometrics, Natural Language Processing, and Disciplinary Analysis

This work uses a base assumption from bibliometric mapping [8], that a research field can be described by a list of important keywords. Previous bibliometrics work ranges from co-citation analysis [9] and examination of conferences [10] to geospatial visualisation of collaboration [11]. Previous work has typically used author assigned key phrases and pre-built domain taxonomies [12], but such resources are not readily accessible here. For this reason we apply an automatic method [13] for extraction of domain terms.

Implicit relations between the extracted topical descriptors can be discovered and described through word co-occurrence analysis, a content analysis technique that was effectively applied to analyse interactions in different scientific fields [8] [12]. This technique was applied to analyse the interconnections between a main field, i.e., fuzzy

logic theory, and other computing techniques [14], a setting that is similar to our analysis of the Internet Science field. A more recent work on co-word analysis [15] outlines several limitations related to the use of keywords and proposes a method to integrate expert knowledge into the process. A main issue with this approach is that it requires a considerable amount of human intervention for the construction of domain specific thesauri. We alleviate this challenge by completely automating the process of identifying topical descriptors and by automatically constructing a domain taxonomy (topical hierarchy).

Internet Science is at the crossroad of domains as diverse as Civic Planning, Psychology and Economics. Each domain has a different level of formality, with a varying number of natural language terms and a more or less deterministic syntax. This has a direct impact on the performance of term extraction tools, with a larger number of correct terms extracted for some domains than for others. The portability of term extraction systems is rarely evaluated across domains, with most studies considering only one domain for evaluation [16] [17]. In [18], different term extraction approaches are evaluated over two domains, a biology corpus and a small general knowledge corpus of Wikipedia articles; term extraction performance is shown to vary depending on the domain. More recent work [19] studies the performance of term extraction systems over three domains (Computer Science, Biomedicine, and Food and Agriculture). That work showed that Saffron, our NLP tool, produces stable results across different domains, and for that reason we use that approach in our work studying Web Science.

This paper builds on previous work conducted by the authors in the related domain of Web Science. A WebSci'12 paper presented initial work in this area [20], and was built upon in a WebSci'13 paper [6] that analysed almost 500 articles (compared to 69 in 2012) and used a small expert survey to (1) aid our interpretations of graph structures and taxonomies and (2) dig deeper when distinguishing between disciplines. In this paper, we use those foundations as a starting point for applying the same approach in this related but distinct domain, in addition describing the practical use of the results in outputs for the Internet Science project, EINS.

## **2.2 Analysis of Internet Science**

Since our previous publication [6], steps include an approach to visualising a given domain's literature based on co-readership [21] and, of particular note here, work to map the an EU FP7-funded Network of Excellence (NoE) [22]. This latter work examined the venues at which members of EINS published, surveying publications to include the word 'Internet' in their title, abstract or keywords; co-authorship networks were also considered. The authors found an imbalance between disciplines identified as predominant in EINS (computer science, physics) and the range of disciplines to do scholarly work about the Internet. This work differs by focusing on what disciplines are present in Internet Science publications, rather than what domains Internet Science experts previously published in.

Other work has explored Internet Science in different ways. At the first Internet Science conference, Dini and Sartori presented an excellent interdisciplinary dialogue (held between a sociologist and an engineer, both active within Internet Science) about the epistemological bases of different disciplines towards developing a method of analysis of the Internet [3]. They conclude that one unified interdisciplinary theoretical framework cannot be attained, but that it is both possible and deeply worthwhile for people with differing disciplinary perspective to collaborate towards a common goal. This is consistent with findings at a recent workshop held on overcoming barriers to interdisciplinary collaboration [5].

### **3 Disciplinary Analysis Method**

In brief, the method involves: gathering a corpus of data; conducting Natural Language Processing to extract topics; a graph analysis and visualisation of the extracted topics; and conducting a ‘community’ analysis of that output.

#### **3.1 Data Gathering Method**

The corpus consisted of the set of papers downloadable from the Internet Science bibliography, online at <http://www.internet-science.eu/biblio>. At the time of the download, August 2014, a total of 208 papers were available. The text-processing tool we use requires that files be in PDF or TXT format. Although some of the papers at the bibliography website were not in a suitable format, most of these papers could be found using Google Scholar. In total, 207 of the 208 papers were processed.

#### **3.2 Natural Language Processing Method**

We processed the files using Saffron<sup>1</sup>, a knowledge extraction framework for understanding research communities [23]. Saffron uses information extracted from unstructured documents using Natural Language Processing techniques [24]. An automatically constructed domain model based on the corpus of data was used, with the approach described in [19]: this domain model was used to construct several linguistic patterns that identify candidate terms based on their context. Noun phrases of maximum 5 words extracted in this way were considered as candidate terms and then ranked based on their length, frequency and embeddedness.

As described in Section 3.1, 207 files were included in the analysis. The Saffron analysis yielded 29,016 phrases that were identified as research term candidates, with an average of 140 candidates per document. The extracted terms were further used for a manual analysis of the corpus and to automatically construct a topical hierarchy. Limitations related to graph visualisation and time constraints mean our analysis considers the best ranked 1000 terms.

---

<sup>1</sup> <http://saffron.insight-centre.org/>

The research terms are not manually curated and therefore include incorrect terms such as ‘research mobility’, which is not an Internet Science research term. Like any other tool, term extraction and analysis has some limitations, and the appearance of ‘research mobility’ as an important term exemplifies the issue of incorrectly extracted terms.

The index used in co-word analysis to measure the strength of relationships between two research terms is defined as:

$$I_{ij} = D_{ij} / (D_i D_j)$$

where  $D_i$  is number of articles that mention the term  $T_i$  in our corpus,  $D_j$  is number of articles that mention the term  $T_j$ , and  $D_{ij}$  is the number of documents in which both terms appear.

Edges are added in the research terms graph for all the pairs that appear together in at least 3 documents. Saffron uses a generality measure to direct edges from generic concepts to more specific ones. This step results in a highly dense, noisy directed graph that is further trimmed using an optimal branching algorithm. An optimal branching is a rooted tree where every node but the root has in-degree 1, and that has a maximum overall weight. This algorithm was successfully applied for the construction of domain taxonomies in [25]. This yields a tree structure where the root is the most generic term and the leaves are the most specific terms.

### 3.3 Graph Analysis and Visualisation Method

We used a network graph tool, Gephi, to build a graph showing links between each of the 1000 terms: nodes are extracted terms and arcs are papers that link them. This let us visually identify clusters of closely related terms. We used the Force Atlas 2 algorithm to layout the graph with the following parameters: Scaling: 2.0; Edge weight influence: 0.0. We used betweenness centrality to weight node importance. Betweenness centrality measures the fraction of shortest paths going through a node [26]: a high value indicates that nodes play an important bridging role in a network.

The lines connecting nodes in the graph represented whether two nodes appeared in the same paper together. Lines could be thicker or thinner, according to the number of times the terms appeared together (thinner lines correspond to fewer co-occurrences, while thicker lines indicate greater co-occurrence). Edges were only visualised for pairs that appeared together in at least 3 documents.

### 3.4 Community Analysis Method

Finally, the Louvain method [27] was applied with resolution 15 to detect ‘communities’, clusters of nodes that were more closely linked with one another than with the rest of the graph. We interpreted ‘communities’ as application contexts ranging from technologies (i.e. open hardware) to disciplines (i.e. network science) and topic areas (i.e. energy).



**Table 1.** The 20 terms with highest betweenness centrality

Internet Science term	Internet Science Betweenness centrality	Web Science term [2013]
Internet governance	322	semantic web
Internet Science	310	social media
data set	211	information retrieval
renewable energy	192	social networking site
centrality metrics	192	social science
charging stations	185	search engine
electricity consumption	183	social networking
energy consumption	180	learning network
operator networks	165	web page
parking spots	164	personal learning environment
data collection	162	social interaction
personal data	160	mobile device
data flows	153	future research
hardware designs	147	internet user
centrality indices	147	uniform resource identifier
connected component	144	web science research
access technologies	140	user interface
access ISPs	138	web community
social interaction	138	web application
content delivery	135	linked data principle

## 4.2 Community Analysis Results

The community detection algorithm returned many ‘communities’ (henceforth: application contexts) with a very sparse distribution. We observed a long tail where most communities were poorly connected and constituted only a very small portion of the graph. For this reason, only application contexts representing at least 1% of the graph were considered, a total of 11/60 detected application contexts. Of the remaining 49 communities, which were very small and not linked to a substantial part of the visualised graph, many consisted of only one term.

Table 2 summarises the 11 considered application contexts, showing for each: a given name; its root node (the node with the highest betweenness centrality); the percentage of the graph that it covered; its number of ‘hot terms’ (terms with a betweenness centrality > 100); and its top 5 terms as ranked by betweenness centrality.

Application context names were chosen based on the topics within each application context, with a weighing towards those topics with a high betweenness centrality.

**Table 2.** Summary of the 11 application contexts

Name	Root node	% graph	# hot terms	Top 5 terms
Measurement technologies	data set	13.2	8	data set; operator networks; data collection; access technologies; social interaction
Energy	renewable energy	13	7	renewable energy; charging stations; electricity consumption; energy consumption; sensor networks
Social Networking	opportunistic networks	12.12	1	opportunistic networks; information privacy; social networking; online social networking; information sharing
Internet governance	internet governance	11.2	6	internet governance; personal data; data flows; content delivery; data protection
Network Science	tie strength	9.8	1	tie strength; ego networks; public key; additional information; tie strength evaluations
Parking	parking spots	8.3	5	parking spots; parking search; opportunistic networks using cognitive heuristics; parking space; cognitive heuristics
Network Science II	centrality metrics	7.3	3	centrality metrics; centrality indices; connected component; video systems; content placement
Architecture / design	internet science	6.6	4	internet science; internet architecture; virtual network mapping; network mapping; current internet
Open hardware and infrastructures	hardware designs	5	1	hardware designs; internet design; open hardware; open source hardware; software code
ISPs and Network traffic	access ISPs	4.7	1	access ISPs; transit traffic; content distribution; access ISP; traffic patterns
Research mobility	information collected	3.3	1	information collected; route selection; BGP datasets; research mobility; routing system

## 5 Discussion

This section is divided into four parts. Firstly, we discuss the key terms and application contexts identified in the above analysis. This is followed by a comparison of these results with the results of applying the same method to Web Science data. Thirdly, we describe the application of our results in EINS outputs, before finally evaluating our approach and discussing future work.

### 5.1 Internet Science Terms and Application Contexts

Our analysis yielded a ranked list of Internet Science terms and a set of application contexts. Topics reflected by the topmost terms include:

- Governance (reflected by the term ‘Internet governance’)
- Energy (‘renewable energy’, ‘charging stations’, ‘electricity consumption’, ‘energy consumption’)
- Network science (‘centrality metrics’, ‘centrality indices’)
- Internet Science (‘Internet Science’)

Terms with less clear mappings include: data set; operator networks; parking spots; data collection; personal data; data flows; hardware designs; connected component; access technologies; access ISPs; social interaction; content delivery.

As can be seen, some terms clearly map to a given topic or discipline, as in the terms mapped to network science above. Other terms, in the above paragraph, are applicable in any of a number of disciplines. ‘Social interaction’, for example, is relevant to disciplines including but not limited to communication, human-computer interaction, psychology, and sociology.

In addition to the ranked terms list, the 11 largest application contexts were analysed. These contexts relate to the terms: for example, the community named ‘measurement technologies’ contains a number of the more ambiguous terms above (data set; operator networks; data collection; access technologies; social interaction). This implies that that particular set of terms co-occurred with one another, perhaps in a particular series of papers.

Among the 11 application contexts, ‘network science’ appeared twice, perhaps representing two sets of network science-driven work that happened not to overlap (i.e. addressing different research questions). It is reasonable to consider ‘network science’ as a single application context of Internet Science, leaving 10 application contexts.

It should be noted that the ‘research mobility’ application context was found due to text referring to the Research Mobility aspect of EINS<sup>2</sup>, rather than to a bona fide research topic: this term can also be discarded, leaving 9 application contexts.

The EINS introductory webpage<sup>3</sup> describes the focus of EINS as focused on “network engineering, computation, complexity, networking, security, mathematics, physics, sociology, game theory, economics, political sciences, humanities, and law, as well as other relevant social and life sciences.” The above findings corroborate part of

---

<sup>2</sup> <http://www.internet-science.eu/mobility>

<sup>3</sup> <http://www.internet-science.eu/network-excellence-internet-science>

that statement (i.e. terms about energy and network science map clearly with areas including engineering, computation and networking), but do not provide strong evidence at this time for presence from, for example, sociology, game theory, economics, political science, humanities, or law. It is possible that NLP may be less good at showing results relating to disciplines which use more varied words (compared with arguably more repeatable terminology in technology fields), and it would be premature to conclude whether fields such as the humanities are definitely under-represented in Internet Science or simply not highlighted by the NLP-led analysis.

## 5.2 Comparing Internet Science with Web Science

The precise delineation between Internet Science and Web Science is not always clear. One might argue that Web Science is a subset of Internet Science (because web technologies are a subset of Internet technologies), while at a 2013 Internet Science / Web Science workshop [28] it was argued that one could view Internet Science as a subset of Web Science. It is beyond the scope of this paper to analyse the relationship between these fields, but nonetheless of interest to provide an initial comparison of the results of applying disciplinary analysis to the two fields.

Table 1 shows the top-ranked terms from an analysis of Web Science papers alongside the top-ranked terms identified from Internet Science in this paper. What is most striking is the almost complete lack of overlap between the two lists of 20 terms: one single term, ‘social interaction’, is the only one to appear on both lists. Furthermore, among the two remaining sets of 19 terms, there are few thematic connections: the Internet Science terms, as discussed above, typically concern topics such as governance, energy and network science, while the Web Science terms centre on topics such as semantic web technologies, social media and information retrieval.

We may also consider the application contexts identified in Internet Science and Web Science. Although initial analysis revealed 11 Internet Science and 8 Web Science contexts, consideration of irrelevant terms such as ‘research mobility’ and synonymous contexts left a total of 9 Internet Science and 4 Web Science contexts, shown in Table 3.

**Table 3.** Internet Science and Web Science application contexts

Internet Science application contexts	Web Science application contexts [2013]
Measurement technologies	Information retrieval
Energy	Personalised learning / elearning
Social networking	Semantic web
Internet governance	Social networking
Network Science	
Parking	
Architecture / design	
Open hardware and infrastructures	
ISPs and Network traffic	

This suggests that even if Internet Science and Web Science do not have fundamental differences and overlap substantially, the two communities that were analysed here and in the 2013 Web Science work [6] do not overlap. Given that, at minimum,

the two communities share similar challenges with respect to overcoming barriers of interdisciplinarity, organisations silos and the like, a greater flow of collaboration between the two is surely desirable.

We note a difference in the number of communities first detected when running the Louvain algorithm on the two datasets. In Web Science, a total of 9 contexts were initially detected, while in Internet Science, 60 were found. There is no significance to this result, which is merely due to one of the parameters of the algorithm.

We also note a difference in the number of candidate terms in each dataset. The Web Science corpus yielded 5361 candidate terms (54 per document), compared with 29,016 candidate terms (140 per document) here. This is not due to a difference between Internet Science and Web Science, but from settings in the algorithm: the first analysis used a manually defined domain model (i.e. domain-specific lexicon) for Computer Science. The analysis presented in this paper used an automatically constructed domain model derived from the corpus itself, which is more specific and adapted to the domain, hence increasing the number and coverage of terms.

### 5.3 Application of Results

A key EINS output within Emergence Theories and Design is an online repository of Internet Science design methods<sup>4</sup>. The results of the Internet Science disciplinary analysis were used to structure the repository, which is a wiki containing a list of design methods. Figure 2 illustrates the front page of the repository.



Fig. 2. Screenshot of the EINS repository of design methods

<sup>4</sup> [http://wiki.internet-science.eu/index.php/Repository\\_of\\_design\\_methods](http://wiki.internet-science.eu/index.php/Repository_of_design_methods)

At the time of writing, the repository contains 49 methods from a diversity of disciplines. A key usability goal is to make the methods accessible by classifying them in useful ways. Application contexts are a key component in that strategy, offering an ideal lens by which to structure (and help users understand) the methods.

Each repository entry consists of: a title; a free text description; a table of relevant links; a table of relevant publications; a list of applicable disciplines; a list of application contexts; any tags; a contributor. Figure 3 shows part of an entry, focusing on the later, more structured parts of the entry.

The screenshot displays a structured entry in the EINS repository. It features five main sections, each with an [edit] link:

- Links**: A table with two columns: Link and Description. The link is [http://www.icsid.org/news/year/2006\\_news/articles267.htm](http://www.icsid.org/news/year/2006_news/articles267.htm) and the description is "A news article on the use of empathic design for baby bottles."
- Publications**: A table with two columns: Publication and Description. The publication is "Leonard, D. and Rayport, J.F., 'Spark Innovation Through Empathic Design', Harvard Business Review, Nov-Dec 1997" and the description is "A key paper on empathic design."
- Disciplines**: A list containing "Design, Engineering, Ethnography, Anthropology".
- Application contexts**: A list containing "Social networking".
- Tags**: A list containing "Design".

**Fig. 3.** Screenshot of part of an entry in the EINS repository of design methods

As can be seen, the method shown in Figure 3 is relevant to the application context of social networking.

Use of application methods to structure the repository offered an additional benefit: the ability to analyse the distribution of repository entries according to application context, and hence to identify under-represented areas. For more on the use of disciplinary analysis to structure and analyse the repository, see [29].

## 5.4 Evaluation and Future Work

As noted in Section 3.2, the terms identified by the NLP analysis are not manually curated and therefore include incorrect terms such as 'research mobility'. This exemplifies a typical limitation of approaches such as NLP, where powerful computation is used to rank terms based on factors such as frequency and co-occurrence, but there is no understanding of term meaning (and hence relevance) to the topic in question.

Section 5.1 observed that some terms clearly map to a discipline (i.e. centrality metrics with network science), but others (such as social interaction) do not. Gaining

more robust insights into the link between terms and disciplines would enhance the insights gained thus far; an expert survey offers an ideal way to complement the strengths of NLP and gain such insights.

Getting input on the term-discipline link from a wide variety of experts overcomes the issues of individual assessment of such linkages, which apply particularly strongly in the context of an interdisciplinary field such as Internet Science. The issue is that two or three researchers, even ones from different fields, cannot represent sufficient disciplinary diversity to analyse and understand the implications of a set of terms sourced from highly interdisciplinary materials.

An example of applying a small survey to bolster the results of disciplinary analysis can be found in [6]: domain experts were asked for their views on what disciplines were most closely related to the top-ranked terms from the dataset. We hope to conduct a similar survey of Internet Science experts in the near future: the upcoming Internet Science conference offers an ideal opportunity to conduct this survey.

An additional benefit from the survey relates to the point in Section 5.1 that NLP may be less good at showing results from disciplines that use more varied words (such as the humanities). Internet Science survey results, combined with existing survey results from Web Science, will let us clarify whether less technologically-oriented disciplines are under-represented in Internet Science or simply less well detected by the NLP-led method used to date, providing insight not only into Internet Science but also into NLP itself.

Concerning other paths for future work, this paper has already presented a brief comparison of results from Internet Science and Web Science. Clarification of the boundaries between Web Science and Internet Science has been identified as an important area for future research [30] [31], and a full NLP-led comparison of Internet Science and Web Science materials represents a way forward in this area.

Finally, this work raises future questions regarding the use of NLP to understand differences in disciplines according to context: for example, how might sociology appear in a Internet Science corpus compared to within a pure sociology corpus? Issues include gaining datasets that are representative of a given domain: for example, the BAWE dataset at Insight has only 111 sociology documents, many of which are short student essays.

## 5 Conclusions

We have described the application of techniques from NLP and network analysis to gain insight into disciplinary representation within Internet Science, revealing a focus on governance, energy and network science (from terms) as well as social networking, parking, and open hardware (from application contexts). We found evidence for the presence of disciplines such as network engineering, computation and networking, and plan further analysis to investigate the presence of less technology-oriented disciplines such as sociology, humanities and law. This supplementary analysis, to be implemented via an expert survey, will also provide insight into the efficacy of NLP when working with non-technology disciplines.

In addition to providing an initial disciplinary analysis of Internet Science, we have demonstrated the applicability of our results for structuring a repository of design methods according to application contexts. We have also described a comparison of the Internet Science results with results from a disciplinary analysis of Web Science, finding a surprising lack of overlap between the two analyses, suggesting that the two communities do not strongly overlap in their current research activities.

## 6 Acknowledgements

The research leading to these results received funding in part from the EU FP7 EINS under grant agreement No 288021 and from the Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight). The authors gratefully thank the survey participants for their time and energy

## References

1. Network of Excellence in Internet Science, Grant agreement 288021, Description of Work, 2011-12-06.
2. Hooper, C.J. and Dix, A. (2013) Web science and human-computer interaction: forming a mutually supportive relationship. *ACM Interactions* 20, 3 (May 2013), 52-57.
3. Dini, P., & Sartori, L. (2013) Science as Social Construction. *International Conference on Internet Science*, 42-51.
4. Guevara, K., & Blackwell, A. (2013) A Reflective Examination of a Process for Innovation and Collaboration in Internet Science. *International Conference on Internet Science*, 37-46.
5. Hooper, C.J., Millard, D.E., Azman, A. (2014) Interdisciplinary Coups to Calamities (workshop) In: Web Science 2014, Bloomington, Indiana, USA.
6. Hooper, C.J., Bordea, G. and Buitelaar, P. (2013) Web Science and the Two (Hundred) Cultures: Representation of Disciplines Publishing in Web Science. *ACM Web Science*, Paris, France.
7. Lungeanu, A., (2014). Understanding the assembly of interdisciplinary teams. *Journal of informetrics*, 8(1), 59-70.
8. Callon, M., Cortial, J., Turner W. and Bauin, S. (1983) From Translations To Problematic Networks – An Introduction To Co-Word Analysis. *Social science information*, vol. 22, no. 2, pp. 191-235.
9. Chen, C. and Carr, L. (1999) Trailblazing the literature of hypertext: author co-citation analysis (1989–1998)," in *10th ACM Conference on Hypertext and Hypermedia*.
10. Henry, N., Goodell, H., Elmqvit, N. and Fekete, J. (2007) 20 Years of 4 HCI Conferences: A Visual Exploration. *International Journal of Human Computer Interaction - Reflections on Human-Computer Interaction*, vol. 23, no. 3, pp. 239-285.
11. Nagel, T., Duval, E. and Heidmann, F. (2011) *Exploring the Geospatial Network of Scientific Collaboration on a Multitouch Table*. 22nd ACM Conference on Hypertext and Hypermedia (demo).
12. Coulter, N., Monarch, I., and Konda, S. (1998) Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, vol. 49, no. 13, p. 1206–1223.

13. Bordea, G. and Buitelaar, P. (2010) DERIUNLP: A context based approach to automatic keyphrase extraction. *Proceedings of the 5th International Workshop on Semantic Evaluation*, Stroudsburg, PA, USA.
14. Lopez-Herrera, A., Cobo, M., Herrera-Viedma, E. and Herrera, F. (2010) A bibliometric study about the reseach based on hybridating the fuzzy logic field and the other computational intelligent techniques: A visual approach. *International Journal of Hybrid Intelligent Systems*, vol. 7, no. 1, pp. 17-32.
15. Wang, Z. Y., Li, G., Li, C. Y., & Li, A. (2012) Research on the semantic-based co-word analysis. *Scientometrics*, vol. 90, no. 3, pp. 855-875.
16. Ananiadou, S. (1994) A methodology for automatic term recognition. *The 15th conference on Computational Linguistics*.
17. Velardi, P. and Sclano, F. (2007) Termextractor: a web application to learn the common terminology of interest groups and research communities. *7<sup>th</sup> Conference on Terminologie et intelligence artificielle*.
18. Zhang, Z., Iria, J., Brewster, C. and Ciravegna, F. (2008) A comparative evaluation of term recognition algorithms. *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, Marrakech, Morocco.
19. Bordea, G., Buitelaar, P. and Polajnar, T. (2013) Domain-independent term extraction through domain modelling. *The 10th International Conference on Terminology and Artificial Intelligence*, Paris, France.
20. Hooper, C., Marie, N. and Kalampokis, E. (2012) Dissecting the Butterfly: Representation of Disciplines Publishing at the Web Science Conference Series. *Proc. ACM WebSci 2012*.
21. Kraker, P., Jack, K., Schloegl, C., Trattner, C. and Lindstaedt, S. (2013) Head Start: Improving Academic Literature Search with Overview Visualizations based on Readership Statistics. *Proc ACM WebSci 2013*.
22. Sahal, A., Wyatt, S., Passi, S., and Scharnhorst, A. (2013) Mapping EINS--An exercise in mapping the Network of Excellence in Internet Science. *The 1st International Conference on Internet Science*.
23. Monaghan, F., Bordea, G., Samp, K. and Buitelaar, P. (2010) Exploring Your Research: Sprinkling some Saffron on Semantic Web Dog Food. *Semantic Web Challenge at the International Semantic Web Conference*.
24. Bordea, G. (2013). *Domain adaptive extraction of topical hierarchies for Expertise Mining* (Doctoral dissertation).
25. Navigli, R., Velardi, N., and Faralli, S. (2011) A graph-based algorithm for inducing lexical taxonomies from scratch. *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*.
26. Barthélemy, M. (2004) Betweenness centrality in large complex networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 38, no. 4, pp. 163-168.
27. Blondel, V.D., Guillaume, J-L., Lambiotte, R. and Lefebvre, E. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008.
28. Fdida, S., Tiropanis, T., Brown, I., Marsden, C., and Salamatian, K. (2013) 1st International Workshop on Internet Science and Web Science Synergies. In: *ACM Web Science 2013*, Paris, France.
29. Hooper, C.J. (editor) (2014) EINS Deliverable 2.1.2: Repository of methodologies, design tools and use cases. Network of Excellence in Internet Science FP7-288021, 2014.
30. Hooper, C.J., Hedge, N., Hutchison, D., Papadimitriou, D., Passarella, A., Sourlas, V., and Wuchner, P. (2014) EINS Deliverable 2.3: Whitepaper on recommendations for funding agencies. Network of Excellence in Internet Science FP7-288021, 2014.
31. Stavrakakis, I. and Hutchison, D. (2014) EINS Deliverable 13.2.1: Internet Science - Going Forward: Internet Science Roadmap (Preliminary Version). Network of Excellence in Internet Science FP7-288021, 2014.