The 1st International Conference on Data-driven Knowledge Discovery: When Data Science Meets Information Science. June 19–22, 2016, Beijing • China

The *1st International Conference on Data-driven Knowledge Discovery: When Data Science Meets Information Science* took place at the National Science Library (NSL), Chinese Academy of Sciences (CAS) in Beijing from June 19 till June 22, 2016. The Conference was opened by NSL Director Xiangyang Huang, who placed the event within the goals of the Library, and lauded the spirit of international collaboration in the area of data science and knowledge discovery. The whole event was an encouraging success with over 370 registered participants and highly enlightening presentations. The Conference was organized by the *Journal of Data and Information Science* (*JDIS*) to bring the Journal to the attention of an international and local audience.

As it is impossible to discuss each presentation one by one, we just mention some highlights and general trends.

Computers

Over the decades, computers have become more and more powerful and as such have had a huge influence on research as they have made investigations possible that were in the past not or hardly possible. Yet, users (scientists, the general public, etc.) still have many demands. Examples of questions that are hard to answer nowadays include: automatically collecting the method(s) used in a study, detecting the main finding(s) of a study (assuming there is one), identifying implicit information, automatically constructing training sets for automated learning, finding all capitalized words in a text, detecting the structure of papers, and automatic function recognition. In addition, merging heterogeneous data has been mentioned, and a possible answer was provided in the sense that using ontologies may provide a step towards a solution.

Big Data and Data Science

What is data science? One possible answer provided by a speaker is: a collection of computational techniques and decision making approaches applicable to massive

amounts of data. It was stated that big data innovated the approach to state governance. It should be mentioned, however, that data for the future are – by definition – not available.

Other colleagues observed that big data encompass among other things: human activities, mobility, and shifts of research interests. The availability of massive data has further led to new methods for data governance and new techniques for decision making.

Data-driven Applications

The Conference provided several examples of practical use of big data: development of agricultural sciences, applications in bio-informatics, and data management for precision medicine being the most striking ones.

Other colleagues discussed data driven decision making, while data facilitating research, for instance, in data-intensive scientific discovery, received quite some attention. In terms of storage of big (and other) data a call was made to create a "commons" such as the National Cancer Institute (NCI) Genomic Data Commons.

When studying data driven applications, one should distinguish between inputs, activities, outputs, and outcomes. Among the outcomes, the following were mentioned: skilled employees, social change, health benefits, policy papers, ecological benefits, and influence on legislation.

When big data are to be used efficiently, cooperation should be the main approach, beating the competitive approach to science.

Metrics – The Structure of Science

Among the metrics-related presentations we mention the "eternal" question of field delineation (what is a field?), and the problem of how to collect relevant data in or about a field. Other aspects discussed during the Conference were: labeling and updating clusters, combining the macro-, meso-, and micro-levels of analysis, and describing the structure of science. It was stated that when trying to find the structure of science we should base on what researchers actually do, not on a theoretical framework. Yet, others emphasized the role of theory.

Only domain experts can answer the question: "How to use available data." Yet, in all practical applications, e.g. when dealing with data related to research evaluation, and derived metrics, these data must be interpreted by experts. Colleagues from the Centre for Science and Technology Studies (CWTS) (Leiden, the Netherlands) presented VOSViewer and CitNetExplorer, two software tools developed for network and data analysis. An example of the use of these tools in the study of the biomarker HER2 was provided by another participant.

The 1st International Conference on Data-driven Knowledge Discovery: When Data Science Meets Information Science. June 19–22, 2016, Beijing • China

Editorial

A presentation on research fronts included the essential question: What is a significant research front? It was suggested that a significant research front is a hot, fast moving research front, with high growth potential. Presenters also paid attention to the issue of modeling.

In a typical example of citation analysis it was noted that only one quarter of all citations are essential. Essential citations are often re-citations in the same publication.

Bridging the relation between metrics and social implications the question "Does more team work lead to more retractions?" was given as food for thought to the audience.

Combining different aspects a participant introduced the notion of convergence. Here convergence is meant in the sense of the coming together of insights and approaches from originally distinct fields. This led to a discussion of the question: What is the relationship between "data driven" discovery and convergence?

Digital Innovation

Points made about digital innovation include:

- Queries can result in too much or too little information;
- How do we find the origin of ideas or of disciplines?
- The path from digital library to digital librarian;
- Innovation pathways;
- Smart cities (characterized among others by many digital innovations);
- Going from systems to cloud services;
- Can digital innovations help to battle data (information?) overload? On this point, it was suggested that constructing representative subsets may be an important step forward.

Science-industry Linkages

Other contributors discussed the assessment of indicators for the economic impact of universities, science-industry linkages in patents, and various types of non-patent linkages besides scientific articles, such as books, handbooks, webpages, and reports. Also the following topics received due attention: tech mining, data mining, foresight, and technology roadmapping.

An interesting observation made in this context was the difference between a professional referent (patent examiner) and a non-professional referent (the scientist when drawing a reference list for an article).

Networks

When studying networks, a multi-layer approach should be the preferred approach in many cases. One should, moreover, include indirect connections provided by neighbors of neighbors, etc.

Another contribution emphasized how to find key nodes in collaborative networks: using amplitude and intensity.

Next to informetrics, bibliometrics, and altmetrics, attention was given to the notion of entitymetrics, referring to the measurement of the impact of knowledge units. Such entities are embedded units in knowledge databases. Network features of such entities may lead to the detection of outstanding interactions between these entities.

How did scientists move from the semantic Web to the knowledge Web?

Finally, it was correctly observed that, when it comes to practical social networks and social media, there are no costs associated with use and non-use of social media. So how can brand loyalty be enhanced? The example of KAIXIN001 provided a case in China where, obviously, brand loyalty was very low.

Social Aspects

Information and data operate among and between humans. Hence they involve social aspects of different kinds. Among these, the following were explicitly mentioned:

- Social aspects *vs*. technical aspects: knowledge science *vs*. knowledge engineers;
- Any technology has some weakness, which can be exploited by those looking for it (cyber criminals);
- Recommendations;
- It was stated that "Everyone you meet knows something you do not know," illustrating the importance of life-long learning;
- Data are not (just) a road to commercialization;
- Importance of data science to society;
- The network society is a society of data: big data and big noises;
- Data handling involves ethical issues, e.g. for those dealing with medical and other sensitive data.
- Retractions *vs*. integrity.

Conclusion

In short, participants of the Conference emphasized the convergence of data science, computer science, and information science, enabling data-driven knowledge discovery to support research, learning, governance, and social and economic development in a big data environment. Yet, they also placed this in the context of future libraries.

The JDIS Editors