Preface

In 1989 the World Wide Web was born in the CERN laboratory, a large and international scientific facility. It was not by chance that this information publishing technology appeared out of a scientific environment because researchers always have needed access to high-quality information about new advances as well as information on the state of the art of their disciplines. At the same time, the scientific community demands the sharing of media that help the rapid dissemination of results and discussion. Today, the practice of science would be impossible without a strong information system that allows the diffusion and storage of all the necessary knowledge. In this context, the Internet and the web represent an authentic revolution, comparable to the birth of the printing press, in the way in which scientists communicate among themselves. This new web era is also bringing about an explosion of new academic results that recognize the diversity of scholarly activity. But perhaps the most important change is the questioning of the earlier way of thinking typified by print publishing. The new potential of the web, participative and open, is favouring the break-up of publishing barriers when it comes to disseminating results. Thus, the open access movement, with the proliferation of digital libraries and repositories as well as the consolidation of free electronic journals, fights for a total democratization of science, bypassing intermediaries and reducing the cost of scientific publishing.

On the other hand, Web 2.0 has made possible a Science 2.0 where social networking sites have emerged that enhance the discussion and sharing of ideas, the comparison of curricula and the popularization of results. In this way, scientists are discovering a huge spectrum of utilities that expand the diffusion of results as well as improve their relationships with other partners. Alongside this flowering of new publishing forms, the web is also bringing about an important change in the evaluation of research. Webometrics – studying the use of hyperlinks as a reflection of citations, and of academic web pages as synonymous with research

production – and currently Altmetrics – analysing the parallelism between citation impact and social visibility – are symptoms of the need for a new way to evaluate this changeable behaviour in research performance, breaking the borders of bibliometric impact and classical publication in print journals.

In this new scenario, specialized search engines for scientists have been emerging over the past few years as a means of gathering this new academic production, accessible via the web. Although academic search engines are indebted to classical citation indexes, this new publishing and communication panorama is changing the appearance of these search services, incorporating heterogeneous research results, adding information sharing tools, and offering a new perception on current scientific activity. In this way, these engines are springing up as web alternatives to the classical citation databases, mainly due to their free access, comprehensiveness and profiling applications.

In the face of such groundbreaking developments, this book aims to present a general review of the most important academic search engines in the market with the intention of making available a brief guide to how these platforms operate, to what and how many materials they index, and to which functionalities they support for a satisfactory search experience. This review has been carried out along a critical path discussing the suitability of these engines for the scientific community, their reliability for evaluation purposes, and the advantages and limitations of each service in meeting the specific and rigorous needs of the researchers. Further, this examination has been carried out via a precise analysis of every section and function, in which each searching feature has been explored and the sources that feed each tool have been tracked.

As the book's subtitle suggests, a quantitative approach is followed in this analytical review, with the aims of measuring the coverage and quantifying the weight of each source. However, the data obtained leads me to suggest an innovative methodology based on the profuse use of crawlers and harvesting procedures which draw a detailed picture of each search service. To some extent, it is an attempt to design an inverse engineering method to capture the essence of the data stored in those services – motivated by the fact that many search engines do not incorporate suitable search interfaces that allow users to take a global view on the coverage of the service as well as providing enough information about their functioning. Therefore, in many cases automatic queries were launched to secure the most detailed coverage possible of the search engine. Another important function of the crawling process is

that it makes possible the detection of duplicated records and misleading information, and the absence of data on authors and documents. Thus, this work tries to present an objective analysis of these search tools, principally based on the extracted data.

Throughout the process of analysis of these products, it could be seen that many of them fulfil various concepts of what academic search engines are. My aim is to highlight the numerous forms that these services may take, signalling that there is no clear definition of what elements are essential in actually designating an academic search engine. The compilation of this list is driven by the popularity and the importance of these developments in shaping the concept of the academic search engine, although each element can differ markedly to the next. In this way, the intention is to present a colourful range of academic search solutions that illustrate different approaches to building data sets, designing search interfaces or structuring contents. On the other hand, the selection of these services has also been shaped by the fact that they share basic elements that reflect the principal characteristics of an academic search engine, helping to flesh out the meaning of the concept.

Addressing a topic in book form presents numerous problems, but publishing a work via web services is inevitably accompanied by the obsolescence of the results. Data on the size of search engines in terms of number of documents indexed, or on the disappearance/appearance of new functionalities, can change dramatically in a short period of time, even if a more quantitative approach is taken, where the data sets are the most unstable elements. For example, although all the figures in this book were updated during February and March 2014, it is possible that during the publication process many aspects will disappear or change. Thinking positively, this should be interpreted as a sign of the dynamic and evolving nature of scholarly engines.

For the above reason, I do not expect to have built a static picture of each search engine, but I hope that this quantitative approximation makes it possible to describe the fundamental characteristics and functionalities of each service, as well as to suggest their advantages and disadvantages for scientific users. An example of the problem of obsolescence mentioned above can be seen in the case of Scirus, which was closed down during the preparation of this book. Despite this, the analysis carried out on that product has been included in the book – first, as a final tribute to its contribution; second, because this could be the last analysis carried out on the site; and, third, because it is a good opportunity to explain for future reference how that engine worked.