Chapter 8
A Unified Approach for Taxonomy-Based Technology Forecasting

Andreas Henschel
Masdar Institute of Science and Technology, UAE

Erik Casagrande
Masdar Institute of Science and Technology, UAE

Wei Lee Woon
Masdar Institute of Science and Technology, UAE

Isam Janajreh
Masdar Institute of Science and Technology, UAE

Stuart Madnick
Massachusetts Institute of Technology, USA

ABSTRACT

For decision makers and researchers working in a technical domain, understanding the state of their area of interest is of the highest importance. For this reason, we consider in this chapter, a novel framework for Web-based technology forecasting using bibliometrics (i.e. the analysis of information from trends and patterns of scientific publications). The proposed framework consists of a few conceptual stages based on a data acquisition process from bibliographic online repositories: extraction of domain-relevant keywords, the generation of taxonomy of the research field of interests and the development of early growth indicators which helps to find interesting technologies in their first phase of development. To provide a concrete application domain for developing and testing our tools, we conducted a case study in the field of renewable energy and in particular one of its subfields: Waste-to-Energy (W2E). The results on this particular research domain confirm the benefit of our approach.

DOI: 10.4018/978-1-61350-038-5.ch008

INTRODUCTION

Any given research field is composed of many subfields and underlying technologies which are related in intricate ways. A solid understanding of how these subfields are linked together as well as how important the different regions of this research landscape are will confer a significant competitive advantage. Currently, information regarding past and current research is available from a variety of channels, providing a rich source of data with which effective research strategies may be formed. These two important trends strongly motivate the development of computational tools for exploiting this data: firstly, the proliferation in technical and academic publications has greatly increased the rate at which relevant knowledge and data are produced and disseminated; secondly, access to this information is constantly improving thanks to the advances in the technologies underlying the web.

Motivation

In order to clarify the intended use of our system, it must be stressed that we are not using "forecasting" in the sense of weather forecasting, where future outcomes are predicted with a reasonably high degree of certainty. It is also important to note that certain tasks remain better suited to human experts. For example, where a particular technology of interest has already been chosen, we believe that a traditional literature review would prove superior to an automated approach. Instead, the proposed framework targets the preliminary stages of technology management, where breadth rather than depth is emphasized. The main focus of our system is on analyzing broad trends occurring in a very large number of documents or other textual sources. By scanning and digesting large amounts of information, promising but less obvious developments can be detected and subsequently brought to the attention of a human expert. This way we capitalize on the strength of computational approaches before making more efficient use of valuable expert time in the critical latter stages of the decision making process.

Knowledge that facilitates forecasting the likely growth and consequences of emergent technologies is essential for well-informed technology management, which is currently relying largely on expert opinion. However, expert decisions can be influenced by personal perspectives or biases. Moreover, acquiring and analyzing such knowledge is hampered by the vast amount of data available in publications. Consequently, sifting through the—often electronically—available R&D literature is time consuming, yet non-exhaustive and subjective. In order to cope with this problem, automated forecasting techniques have been developed in recent years (see Background section). A remaining challenge is related to the knowledge organization of the acquired data. For example, in order to elucidate the advances of technologies, we want to answer questions like: "How many scientific articles have been published in peer-reviewed journals on the topic of solar energy recently?" Intelligent search techniques capable of grouping semantically similar concepts are therefore needed, such that the term "parabolic trough" is subsumed under solar energy related technologies and hence articles about it should be included in the analysis. This underlying challenge of managing and structuring the vast amount of available knowledge from web resources is similar in web-based Technology Forecasting and general Semantic Web applications. However the former has yet to fully benefit from the advances of the latter. In particular, the state-of-the-art Technology Forecasting tools hardly make use of ontologies or taxonomies, the standard form of knowledge representation for the Semantic Web (Shadbolt, 2006).

The major novel aspect in the presented work is a modular and automated approach, which streamlines data acquisition, keyword extraction and taxonomy creation as the basis for trend detection. The framework provides evidence for