Peaks, Slopes, Canyons, Plateaus: Identifying Technology Trends throughout the Life Cycle

Irina V. Efimenko[*], Vladimir F. Khoroshevsky[**] and OlegV. Ena[*]

[*] *{iefimenko, ovena}@hse.ru*
Institute for Statistical Studies and Economics of Knowledge, Higher School of Economics (Russia)

[**] *v.khor@mail.ru*
Dorodnicyn Computing Centre of Russian Academy of Sciences (Russia)

Introduction

Automated identification of new technology trends (trend monitoring, trend hunting, trend watch) is among the hot topics in technology management. Despite many beneficial results in this field, almost no solutions allow users to escape from getting too general or garbage results which make it impossible to identify trends at the stage of weak signals. Lack of attention is paid to automated labeling and merging (for the 'same' trends).

Our approach aimed at overcoming such drawbacks is based on the 'BlackBox' principle. The concept of a technology trend (TT) is characterized by a complex nature, low formalization level, blurred boundaries, and high degree of domain dependency leading to the need for expert knowledge. For all that, 'Big Data' in IT and 'Genome Editing' in Healthcare should have some similar features which actually allow us to name both phenomena 'a TT'. This leads us to an idea of hunting for domain independent 'external signs' (trend indicators) while letting a TT itself stay a black box for an observer.

We employ Gartner's Hype Cycle in our methodology. We build an elaborate ontology of a TT and a system of indicators of TTs 'presence' in documents of various genres. The indicators are interrelated with the ontology through linguistic and extra linguistic markers. Both markers and text genres are mapped onto the phases of a technology life cycle. The ontology-driven information extraction (IE) is carried out.

Basic Hypotheses and Assumptions

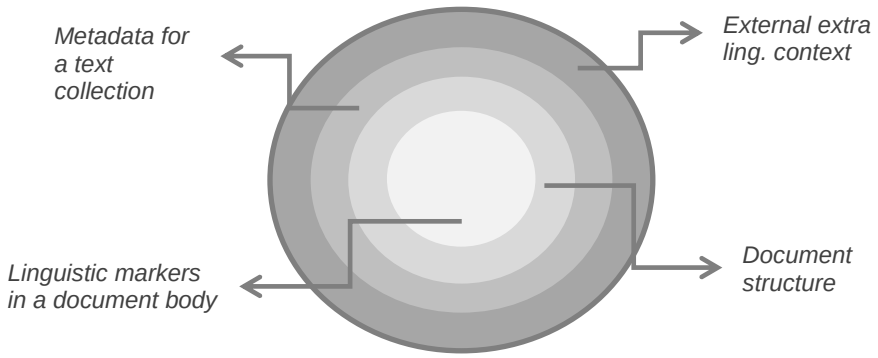
It is well known that each Hype Cycle[1] drills down into the five key phases of a technology life cycle: Technology Trigger, Peak of Inflated Expectations, Through of Disillusionment, Slope of Enlightenment, Plateau of Productivity. We use the following basic hypotheses:
1. Specific (extra) linguistic markers should be used for each phase; 2. TT monitoring should be carried out based on a mixture of text genres. However, each phase has its own most relevant genres; 3. Patent analysis is most important for the phases of Slope of Enlightenment and Plateau of Productivity; 4. Surge and loss of interest could be fixed based on processing of technology news and specific types of foresight and analytical reports. It is relevant for the Peak of Inflated Expectations and the Through of Disillusionment; 5. Scientific papers and R&D reports are relevant for the Technology Trigger. It seems to be the most interesting phase for the TT monitoring task (at the stage of weak signals).

BlackBox Principle: Heuristics and Indicators

The black box approach is related to deriving knowledge on TTs from trend indicators. All heuristics based on indicators correspond to one of the levels represented in Figure 1.

Figure 1: Types of indicators

[1] http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

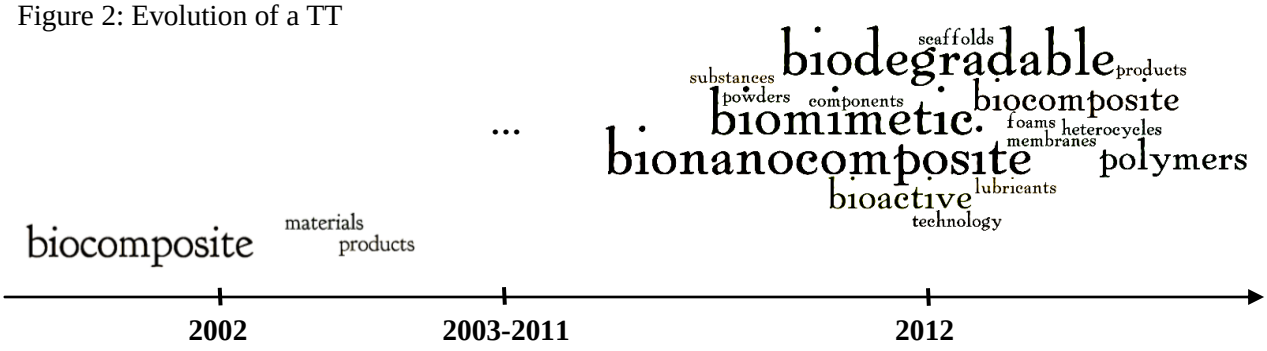


Most of indicators are either an event or a linguistic marker. Thus, **linguistic scales** appearing in texts, mostly in scientific papers, are exemplary used to identify a TT in early stages of its evolution. They correspond to (a) new / existing needs (customers' pains); (b) resources, values; (c) S&T issues (tasks); (d) technological parameters (the change of). E.g. "For the marine industry, MRE isolators could *greatly* decrease the level of vibrations transmitted from the machines to the shell of the ship and the opposite, resulting to smaller fatigue loads and a *much* more comfortable journey". The fact of forming a new **regulatory committee** is a potential evidence for a more mature TT (the need for regulation means that the phenomenon is newish, but already having significant impact). All indicators are grouped into the semantic fields. These fields all together build an ontology system which is used for the ontology-driven information extraction. Weights are assigned to each indicator or a group of indicators. Some of indicators are genre-specific, others are universal. At the stage of IE, text collections are processed to identify meaningful fragments with indicators. Digests of text collections are formed. Hybrid statistical and linguistic methods are used to extract terms, which describe TTs, from these digests.

Case Study and Discussion

An approach and the system were tested using text collections in Green Energy domain (131,477 documents in English and in Russian, scientific papers, technology news, etc.; 2002-2012). For an example of results for news text collection (mostly got from http://cleanedge.com/) see Figure 2[2].

Figure 2: Evolution of a TT



The results show that the proposed approach not only allows users to get relevant and interpretable information on TTs. It also contributes to development of the theory of TTs and to new heuristics for trend hunting. Thus, 'splitting' of terms can be considered to be a meaningful phenomenon for automated identification of new TTs in various stages. A number of new indicators was built based on the processing results which provided recursive refinement of the algorithms. Further R&D are to be focused on extension of ontologies for enhanced domain-specific versions, deeper examination of relations between indicators and a TT life cycle, and new algorithms for merging, labeling and visualizing TTs.

[2] Wordle (http://www.wordle.net/) was used to visualize clouds of terms.