Computers in Industry 100 (2018) 244–257

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Extracting and mapping industry 4.0 technologies using wikipedia

Filippo Chiarello[a,*], Leonello Trivelli[b], Andrea Bonaccorsi[a], Gualtiero Fantoni[c]

[a] Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Largo Lucio Lazzarino, 2, 56126 Pisa, Italy
[b] Department of Economics and Management, University of Pisa, Via Cosimo Ridolfi, 10, 56124 Pisa, Italy
[c] Department of Mechanical, Nuclear and Production Engineering, University of Pisa, Largo Lucio Lazzarino, 2, 56126 Pisa, Italy

ARTICLE INFO

Keywords:
Industry 4.0
Digital industry
Industrial IoT
Big data
Digital currency
Programming languages
Computing
Embedded systems
IoT
Internet of things

ABSTRACT

The explosion of the interest in the industry 4.0 generated a hype on both academia and business: the former is attracted for the opportunities given by the emergence of such a new field, the latter is pulled by incentives and national investment plans. The Industry 4.0 technological field is not new but it is highly heterogeneous (actually it is the aggregation point of more than 30 different fields of the technology). For this reason, many stakeholders feel uncomfortable since they do not master the whole set of technologies, they manifested a lack of knowledge and problems of communication with other domains.

Actually such problem is twofold, on one side a common vocabulary that helps domain experts to have a mutual understanding is missing Riel et al. [1], on the other side, an overall standardization effort would be beneficial to integrate existing terminologies in a reference architecture for the Industry 4.0 paradigm Smit et al. [2].

One of the basics for solving this issue is the creation of shared semantic for industry 4.0. The paper has an intermediate goal and focuses on the development of an enriched dictionary of Industry 4.0 enabling technologies, with definitions and links between them in order to help the user in actively surfing the new domains by starting from known elements to reach the most far away from his/her background and knowledge.

1. Introduction

Industry 4.0 is getting the center of the scene with respect to the future of production systems in advanced countries and to its economic and social implications. It is considered as the new fundamental paradigm shift in industrial production. The new paradigm is based on the advanced digitalization of factories, the Internet, and future-oriented technologies bringing intelligence in devices, machines, and systems [3]. Despite its growing popularity and the great expectations in terms of innovation impact, the concept of Industry 4.0 remains strongly linked to technologies and frameworks that have been heavily researched and analyzed in the last decades. In particular, Industry 4.0 can be seen as a smart recombination of existing technologies and some new technologies and their application to the manufacturing environment [4]. This recombinant nature has led some authors to claim that it is nothing more than a re-labeling of old technologies, such as Computer Integrated Manufacturing [5].

Yet other authors claim that this new wave of technology is fundamentally different from previous technologies and not just an amalgamation. In order to address the question whether Industry 4.0 is a new paradigm, or rather a re-labeling of existing technologies, a preliminary activity is needed, namely the delineation of the field and the clustering of technologies covered in the perimeter. It turns out that this activity is extremely challenging in the case of Industry 4.0, for a number of reasons we discuss in great detail. Faced with the complexity of Industry 4.0 existing delineation and clustering methodologies can be considered inadequate.

In this paper we develop a novel approach, test it, and show its superior performance with respect to other approaches.

The key features of the approach are as follows:

i) the description of Industry 4.0 is offered in the form of an “enriched dictionary”, or an ordered and comprehensive collection of lemmas, each of which are associated to full scale definitions and descriptions and to explicit linkages to other lemmas;
ii) the description of constituent technologies offered in the enriched dictionary is not obtained from individual experts, but is generated by accessing appropriate pages of the online encyclopedia Wikipedia;
iii) the total number of technologies covered is more than 1200, linked with more than 39,000 semantic relations;
iv) the perimeter of Industry 4.0 is not defined externally to the

* Corresponding author.
E-mail address: filippo.chiarello@destec.unipi.it (F. Chiarello).

https://doi.org/10.1016/j.compind.2018.04.006
Received 25 September 2017; Received in revised form 1 February 2018; Accepted 10 April 2018

technology (by experts, government policies or other external sources) but is generated endogenously by examining the linkages between technologies described in the Wikipedia pages;

v) the update of the descriptions of technologies in the dictionary takes place in real time due to the distributed, parallel and self-controlled activities of authors in the worldwide community of contributors to Wikipedia;

vi) new technologies are automatically included in the dictionary if they exhibit a given threshold of connectivity with those already included in the perimeter.

The paper is structured as follows. We first characterize the field of Industry 4.0 and discuss why it creates a challenge for field delineation and clustering. Second, we review the recent literature on delineation and clustering methods and we show their limitations with respect to the features of Industry 4.0. We then develop a novel methodology to identify and describe in great detail all the technologies involved in the field. A case study shows how the methodology works in operational terms. Finally, the results are discussed and future developments of the research are outlined in the last section of the paper.

2. Industry 4.0 as a multi-technology multi-stakeholder field

Industry 4.0 is the main keyword used by researchers, policy makers and entrepreneurs when describing how worldwide industrial systems will evolve in the near future by leveraging Internet connected technologies to generate new added value for organizations and society [6]. The growing interest is confirmed by the increasing number of academic papers focusing on topics that are related to the so-called “Fourth Industrial Revolution”. As shown in Fig. 1 the query “Industry 4.0” generates 967 papers. Even if the query is very sharp and does not include all the research efforts on the single “enabling technology” it demonstrates an exponential growth of the topic. In Fig. 1 a projection represented by the dotted line is included. The projection has been drawn by considering a constant increase of the derivative calculated as the average of the last 4 years. Our forecast is that in 2017 there will be 575 new papers in Scopus; this estimate is supported by the fact that about 200 papers have been already published before June 2017 (represented by the point). Since previous analyses in Scopus demonstrate a delay between publication and loading of 5–6 months our forecast seems to confirm a growing interest on the topic.

Table 1 shows how the scientific production on Industry 4.0 is divided among the main research fields (multiple attributions are possible in Scopus).

In particular, it is possible to identify field specific-technologies that refers just to one or few sectors/business areas, and general purpose technologies that can be exploited in several sectors/business areas.

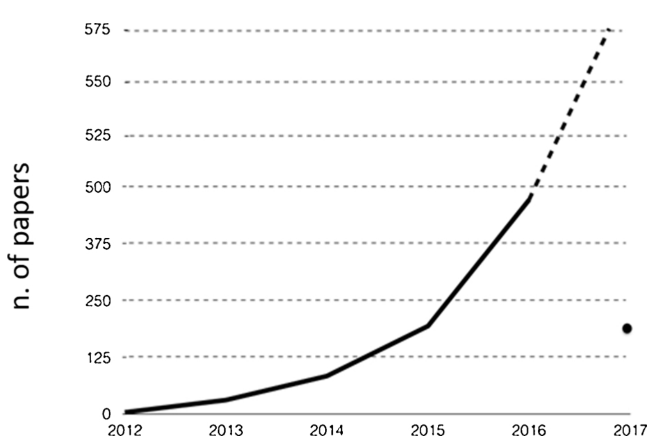

Fig. 1. Trend of publications on Industry 4.0 (Title, Abstract, Keywords). Source: Scopus. Date: 06/06/2017

Table 1
Breakdown of industry 4.0 papers per research field.
Source: SCOPUS date: 06/06/2017

Subject Area	Number of Publications
Engineering	645
Computer Science	410
Business, Management and Accounting	185
Decision Science	134
Material Science	90
Mathematics	87
Chemistry	52
Physics And Astronomy	45
Social Sciences	34
Energy	30

Formulated initially in Germany in 2011, the Industry 4.0 paradigm has been quickly translated, adapted and reinterpreted in developed and developing countries. Table 2 offers a compilation of official documents of governments, agencies and international organizations that in a few years after the initial formulation have embraced the concept.

Despite this rapid and impressive convergence of interest (in itself a clear demonstration of the interdependence of policies across the world), there is no common ground in the definition and delineation of the field even if a first definition of the goal of industry 4.0 have been presented since 1998 [7].

More precisely, while there is a reasonable convergence on the architectural definition of Industry 4.0, as defined in a relatively loose way, there is still considerable disagreement and misalignment with respect to constituent technologies [1,2,8].

Furthermore, many constituent technologies are included in the definition of Industry 4.0, and hence described in these documents, from a variety of perspective that reflect mainly the huge variety of application domains. In other words, technologies are often described not only with respect to their fundamental engineering principles and related dimensions of performance, but with respect to specific applications to various manufacturing or service operations. In these applications the specific working of technologies and the associated dimensions of performance are indeed quite diverse.

Grangel and González (2016) develop a deductive rule-based system able to identify conflicts among AutomationML documents, named ALLIGATOR. It is interesting for the present work to notice how ALLIGATOR has the function to interoperate and align information models between a vast variety of areas (manufacturing, security, logistics) at a micro/plant level. In other words, this paper highlights the fact that one of the main problems of Industry 4.0 is the integration of models and concepts typically developed in their respective domains.

To offer an example of this state of affairs, let us consider the case of RFID (Radio Frequency Identification and Detection) technology. One of the main uses of RFID technology, that is, the detection of the location of a tag moving along a known path with known speed, as illustrated in one of the basic patents [9], can indeed be applied for largely different purposes (safety, tracking, localisation) and in various company areas (production, logistics, maintenance). In practice, each of these applications will develop the basic technology in different directions. Depending on the applications we will find largely different descriptions of the technology involved. Fig. 2 offers a Value Chain-like representation of Industry 4.0, showing the wide range of applications of constituent technologies.

An interesting consequence of this state of affairs is that there is disagreement also at the higher level of government documents describing Industry 4.0 as the main object for innovation and industrial policies: when describing the main components of Industry 4.0 the French government uses 47 technologies, against 39 technologies for the Italian government.

Summing up, the recombinant nature of Industry 4.0 creates several