International Journal of Distributed Systems and Technologies
Volume 11 • Issue 2 • April-June 2020

A Thematic Analysis of the Articles on the Internet of Things in the Web of Science With HAC Approach

Asefeh Asemi, Corvinus University of Budapest, Budapest, Hungary
Fezzeh Ebrahimi, Corvinus University of Budapest, Budapest, Hungary

ABSTRACT

This research was carried out using the bibliometric method to thematically analyze the articles on IoT in the Web of Science with Hierarchical Agglomerative Clustering approach. First, the descriptors of the related articles published from 2002 to 2016 were extracted from WoS, by conducting a keyword search using the "Internet of Things" keyword. Data analysis and clustering were carried out in SPSS, UCINET, and PreMap. The analysis results revealed that the scientific literature published on IoT during the period had grown exponentially, with an approximately 48% growth rate in the last two years of the study period (i.e. 2015 and 2016). After analyzing the themes of the documents, the resulting concepts were classified into twelve clusters. The twelve main clusters included: Privacy and Security, Authentication and Identification, Computing, Standards and Protocols, IoT as a component, Big Data, Architecture, Applied New Techniques in IoT, Application, Connection and Communication Tools, Wireless Network Protocols, and Wireless Sensor Networks.

KEYWORDS

Co-Occurrence Word, Hierarchical Agglomerative Clustering (HAC), Hierarchical Clustering, Internet of Things (IoT), Thematic Analysis, Web of Science (WoS)

INTRODUCTION

The advent of Industry 4.0 and the prevalence of IoT set the scene for the interactions between the objects as well as between the objects and humans. This technology has numerous capabilities such as smart treatment and energy efficiency. "The Internet of Things" (IoT) is a concept that allows an entity (a human, animal, or object) to send data through the sensors attached to it via the communication networks such as the Internet and the Intranet (Weber, 2010)." In IoT, each natural or legal object in the surrounding environment has an ID and these objects communicate with one another in an integrated medium. IoT, as one of the most important new technologies, has influenced all dimensions of human life. On the other hand, the scientific and research activities in the area of IoT have escalated drastically in the past several decades. The increase in the number of studies on IoT has led to an increased understanding of this concept and the increasing use of this technology in different sectors and industries. Moreover, given the increased flow of information and the importance of using IoT, it is substantially important to identify the hidden themes pertinent to IoT and to cluster those themes. This analytical approach can help find the hidden relationships in a field of science, develop a concept in a certain time period (Assefa & Rorissa, 2013), identify the trends in a certain field (Wang & Inaba, 2009), identify the important themes in a subject area, and discover the topics

DOI: 10.4018/IJDST.2020040101

of interest and predominant notions in the works of researchers (Kumar & Jan, 2012) In Gartner's hype cycle (Gartner Inc), IoT is placed on top of the diagram. This diagram shows the "emerging technologies roadmap." As seen in Figure 1, IoT is on top of this diagram based on a 5- to 10-year estimate.

The results of the investigations by Gartner's research institute (Gartner Inc) suggest that more than 25 billion devices will be connected to the Internet or other communication networks around the globe via IoT-based services. IoT has been classified differently from different aspects. For instance, Zhou classifies IoT into the following four major categories (Figure 2):

- **M2M (machine to machine):** It uses various machines to receive the events via a network connected to a central server so as to convert the collected events into meaningful information;
- **RFID (radio-frequency identification):** It uses radio waves to transmit data to a reader to allow for identification and tracking;
- **WSN (wireless sensor network):** These networks include a set of automated and spatially dispersed sensors that monitor the physical and environmental conditions such as temperature, pressure, motion, and pollution. These sensors transfer their data to a central location via the network;
- **SCADA (supervisory control and data acquisition):** It is an automated system based on the closed-loop control system, a smart system or CPS, which connects, controls, and monitors the equipment via a network.

Numerous articles have been published on this subject. Examples are the articles authored by Munoz (2017), Yang, Huang, Sun & Zhang (2017), Hou, Mao, Zhao & Du (2015), Makizadeh, Hazeri & Keikhaee (2016), Hazeri, Makizadeh & Mobasheri (2016) who discussed the e-Government thematic clustering, NPS pollution, lifecycle assessment, diabetes, addiction, and financial performance index, respectively. These researchers have referred to the common and key themes in the given areas. The

Figure 1. Gartner's hype cycle (www.gartner.com)