A review of the current state of research on water resources management

Claudia Ifrim[1], Marko Harasic[2], Adrian Paschke[2], Florin Pop[1], and Valentin Cristea[1]

[1] Faculty of Automatic Control and Computers
Computer Science Department
University Politehnica of Bucharest, Bucharest, Romania 060042
claudia.ifrim@hpc.pub.ro, florin.pop@cs.pub.ro,
valentin.cristea@cs.pub.ro
http://acs.pub.ro/

[2] Fraunhofer Institute for Open Communication Systems
Berlin, Germany
marko.harasic@fokus.fraunhofer.de, adrian.paschke@fokus.fraunhofer.de

Abstract. The purpose of this paper is to review the interest on water and water resources management in the scientific community in order to understand the current state of research and how it is tagged by the authors. Based on a set of available abstracts and freely available papers extracted using a crawler developed by us based on Scopus API [1] we adapt the data set to match a specific JSON format used by VIPRA, a visual platform for topic mining assisted browsing and searching of large datasets. We will present the results obtained using VIPRA platform and we will also analyze what keywords are assigned by authors to mark their papers in water resources management topic. The first section of this paper will present the objectives of our study, followed by a section where we will describe the methodology used to achieve our goals and one with the outcome of our research. In the final sections we will expose our conclusion based on the analyzed papers and we will present our future plans to extend our study.

Keywords: water resources management · meta data · citation network · hydrology

1 Introduction

Despite the fact that we have a vast amount of water available on the globe, years of unsustainable management resulted in appearance of water shortages crisis points in many regions of the planet. We face water availability problems when the demand for water exceeds the amount available water resources during a certain period. Those types of problems occur frequently in zones with intensive industrial or agricultural activity, low rainfall zones and high population density.

So far, there have been no major incidents in Europe caused by severe water scarcity, but taking into account recent climate change and increasing global water consumption, water resource management has become a topic of interest both politically and academic.

Nowadays, resource management decisions should be based on knowledge and insights gained from environmental information and related data sources. Thus, usability of available information is hampered, the larger the number of data sets is the less structured the data are because we don't have a global standardization for the data sets format that should be respected.

Environmental studies are often spatial and related to certain locations or regions of interest (ROI).

The aim of the current paper is to study how research articles on water resources management topic are described by authors, what keywords are commonly used and how relevant those are if we study the topic trends of those articles.

2 Methodology

By analyzing the data sources we could use for this paper, the available APIs[2] and other methods by which we can build a relevant data set for our analysis, we chose the solution to develop a crawler based on Scopus API.

We built the data set running our custom crawler with the paper entitled "A review of the current state of research on water, energy, and food nexus" as a seed. For each retrieved paper we saved the keywords, title, url, abstract, references, authors (id, name and profile url), publishing date and meta data related to paper as DOI, document type, ISSN, original language, publisher and source type.

Based on the papers referenced in each article we build a citation network.

The final data set used for the analysis reported in this paper has a number of 14951 records obtained by a breadth-first search.

This data set was transformed into our internal JSON format and then imported into our Dynamic LDA tool named VIPRA.

VIPRA is a system that allows searching and browsing through large data collections. It does so by importing data into two databases and running a topic modeling algorithm over the imported data to generate a topic model, a virtual list of numbered topics that cluster documents by similarity. This technique should make it easier for the user to explorer the data.

Topic modeling [3] [4] is a technique from the field of natural language processing. It is used to cluster a collection of documents into one or more abstract groups, so called topics. An algorithm analyzes word frequencies to correlate documents to each other and based on that, calculates topic proportions for these documents. Topic modeling is a type of unsupervised learning, because the algorithm receives no pre-configured topics to assign documents to. Instead it creates its own topics and describes them with proportional relations to documents and word association probabilities.

Trend detection [5] [6] describes a technique that identifies and tracks emerging and submerging topics, called trends, in continuous data streams via several markers, such as keywords or meta data. The objective of trend detection is mainly to analyze fluent data in regard to various aspects, such as advertisement, statistics, opinion tracking and analysis. One method of detecting trends is to scan available keywords for high frequency. The assumption is that if a word suddenly appears a lot more often than usual, a new trend is emerging. Finding and collecting these and similar words in the data stream is the subsequent task, involving natural language processing.

The detection of trends in data, textual or not, is important for the analysis of evolving data. In the context of VIPRA system, trend detection is represented as a part of the dynamic topic modeling approach that detects topic changes in time related data.

3 Research outcome

In order to review the interest on water and water resources management in the scientific community we used the same data set to analyze the trends in topics related to water resources management and we also analyzed what keywords are assigned by authors to mark their papers in water resources management topic.

3.1 Trends in topics analysis

As the training takes some time, it is still unclear, how the results will look like. It is planned to investigate the topics and create furthermore graphs on the keywords here ...

3.2 Keywords analysis

To see how articles dealing with water resource management topics should be framed, we've reviewed a series journals as:

- Advances in Water Resources [10]
- Computers & Geosciences [11]
- Environmental Monitoring and Assessment [12]
- Hydrology and Earth System Sciences [13]
- Journal of Hydrology [14]
- Water [15]

Based on the way these journals are described and from the categorization of their areas of interest, the main categories in which we consider that we can find articles dealing with topics on water resources and water resource management are:

- Water resources management
  - water resources systems

  - surface water resources systems
  - subsurface water resources systems
  - water monitoring
  - remediation and protection of water resources
  - water resources planning
  - adaptive management
  - water demand management
  - national water policy
  - international water policy
- Water and wastewater treatment
  - water purification
  - water reuse
  - constructed wetlands
  - treatment wetlands
- Urban water management
  - urban drainage
  - storm water management
  - local water storage

We will observe the usage of those keywords in our data set.

The data set used for our analysis contains 14951 articles out of which only 5538 have associated keywords. The number of the keywords used by authors to describe their articles is 30122, counting only 15226 unique keywords (see Fig. 1). The most used keywords can be found in Table 1 and visualized in Fig. 2.

Table 1. First ten most used keywords to describe the articles topic.

Keyword	Number of uses
Climate change	279
Biomass	133
Microalgae	130
Irrigation	129
Agriculture	97
Nitrogen	92
Sustainability	83
Co	81
Environment	74
Temperature	74

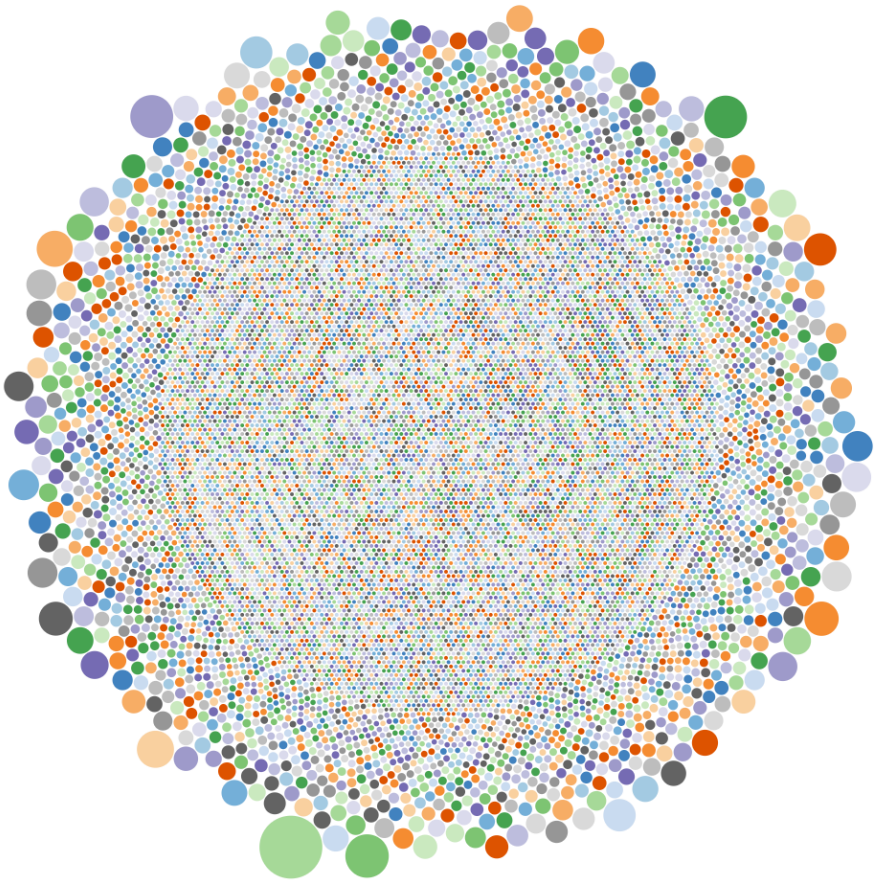

Fig. 1. Graphic representation of all unique keywords associated to articles in our dataset based on their number of appearances.

Table 2. First ten most used keywords containing word "water".

Keyword	Number of uses
Groundwater	60
Virtual water	30
Wastewater	27
Wastewater treatment	18
Groundwater recharge	14
Ground water	14
Integrated water resources management	13
Surface water	11
Groundwater management	11
Drinking water	8

Fig. 2. Graphic representation of most relevant 120 unique keywords associated to articles in our dataset based on their number of appearances.

From the total number of unique keywords only 410 contain the word "water" (see Fig. 3). The most used keywords from this subset can be found in Table 2 and visualized in Fig. 4.

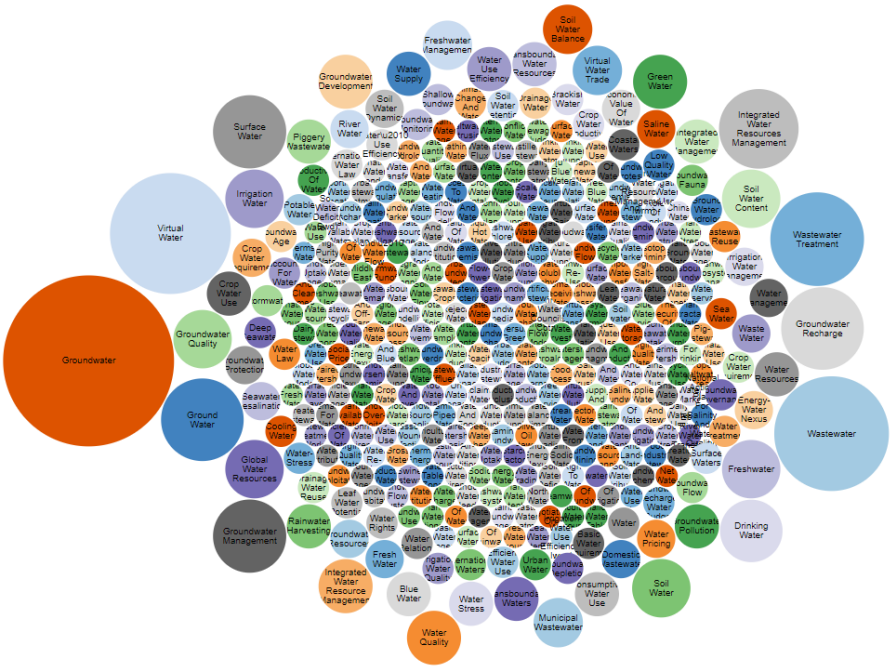

Fig. 3. Graphic representation of unique keywords containing "water" string associated to articles in our dataset based on their number of appearances.

Fig. 4. Graphic representation of most relevant 120 unique keywords containing "water" string associated to articles in our dataset based on their number of appearances.

4 Conclusion

We built a data set using a Scopus API crawler, a set containing 14951 records of extracted articles starting from a search after the article "A review of the current state of research on water, energy, and food nexus".

On this data set we analyzed the trends in topics using our Dynamic LDA tool named VIPRA and we also analyzed the keywords used by authors to describe their articles.

Observing the results of the keyword analysis found in our dataset, we can conclude that the subjective choice of these keywords by the authors of articles makes the area of interest where the article could fit to be more general than the subject the article treats actually. Most articles dealing with water resource management issues are now part of the climate change area of interest.

5 Future work

We plan to extend this study using a larger set of data and enrich it with other meta data such as author affiliation, the country they come from and the full text of the articles where possible.

Based on the new dataset, we will look at the geographical area where interest in this topic has been expressed and we will compare its evolution in time based on the publication data of each article.

References

1. Scopus API Homepage, https://dev.elsevier.com/. Last accessed May 2018
2. APIs for scholarly resources, http://guides.lib.berkeley.edu/information-studies/apis. Last accessed May 2018
3. David M. Blei and John D. Lafferty. Topic models, 2009.
4. David M. Blei. Probabilistic topic models. Commun. ACM, 55(4):7784, April 2012.
5. Jey Han Lau, Nigel Collier, and Timothy Baldwin. On-line trend analysis with topic models: twitter trends detection topic model online. In COLING, pages 15191534, 2012.
6. Hao Zhang, Gunhee Kim, and Eric P Xing. Dynamic topic modeling for monitoring market competition from online text and image data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 14251434. ACM, 2015.
7. ALCAMO, Joseph; FLRKE, Martina; MRKER, Michael. Future long-term changes in global water resources driven by socio-economic and climatic changes. Hydrological Sciences Journal, 2007, 52.2: 247-275.
8. KOSTAS, Bithas; CHRYSOSTOMOS, Stoforos. Estimating urban residential water demand determinants and forecasting water demand for Athens metropolitan area, 2000-2010. South-Eastern Europe Journal of Economics, 2006, 1.1: 47-59.
9. DIXON, A., et al. Measurement and modelling of quality changes in stored untreated grey water. Urban Water, 2000, 1.4: 293-306.
10. Advances in Water Resources, https://www.journals.elsevier.com/advances-in-water-resources/. Last accessed May 2018
11. Computers & Geosciences, https://www.journals.elsevier.com/computers-and-geosciences. Last accessed May 2018
12. Environmental Monitoring and Assessment, https://link.springer.com/journal/10661. Last accessed May 2018
13. Hydrology and Earth System Sciences, https://www.hydrology-and-earth-system-sciences.net/index.html. Last accessed May 2018
14. Journal of Hydrology, https://www.journals.elsevier.com/journal-of-hydrology/. Last accessed May 2018
15. Water, http://www.mdpi.com/journal/water. Last accessed May 2018