What are the Evolution Patterns of Research Strands in Water-related Disease Transmission in Groundwater?

Yiding Zhang[1], Motomu Ibaraki[2], and Franklin W. Schwartz[2]
[1] Environmental Science Graduate Program, The Ohio State University, Columbus, OH 43210, USA zhang.1381@osu.edu
[2] School of Earth Sciences, The Ohio State University, Columbus, OH 43210, USA

Introduction

Groundwater is associated with a significant fraction of all water-related disease outbreaks. The major contaminants from groundwater include microbial pathogens, heavy metals, and organic compounds. The objective of this research is to interpret the patterns of research strands in groundwater related diseases using machine learning, text mining, and visualization techniques.

Methods

- **Article Retrieval**
  - 426 research articles from 1971-2017;
  - Search terms: groundwater & disease.
- **Language Processing**
  - Article abstracts were preprocessed by removing stop words and stemming using a natural language processing tool (NLTK).
- **Article Similarities and Clustering**
  - Article similarities were calculated using cosine similarity;
  - Louvain Method was used in clustering.
- **Topic Extraction**
  - Key information was extracted and scored (weighted) using RAKE, which is a keyword extraction algorithm.

Clustering

- 11 clusters were identified based on calculated article similarities;
- The number of articles in each cluster ranged from 10 to 73.

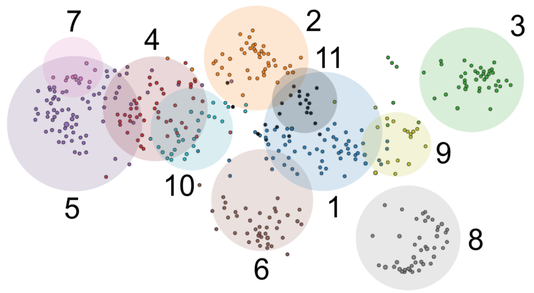

Fig. 1 Clustering based on article similarities

Topic Identification

Cluster No.	No. of articles	Keywords / Major topics
1	58	Bacteria and virus contamination and related disease in drinking water, surface water, and groundwater
2	44	Arsenic enrichment and blackfoot disease
3	45	Bacteria (legionella) in spring water and thermal water
4	45	Chemical exposures
5	73	Chronic arsenic poisoning and cancer
6	40	Drinking water studies and effect of climate change to water-related diseases
7	10	Groundwater problems in Bangladesh
8	43	Waterborne disease outbreaks and case studies
9	17	Soil contamination and leachate
10	28	Fluoride contamination and kidney disease
11	23	Risk assessment

Disease Patterns in the Clusters

- Two cluster groups were generated according to the contaminants-chemicals / heavy metals and pathogens;
- The major diseases are cancer and diarrhea respectively.

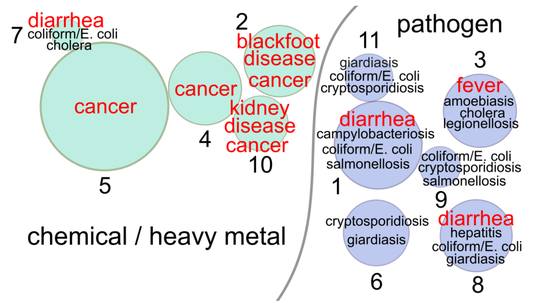

Fig. 2 Major diseases in the clusters

Water Type Classification

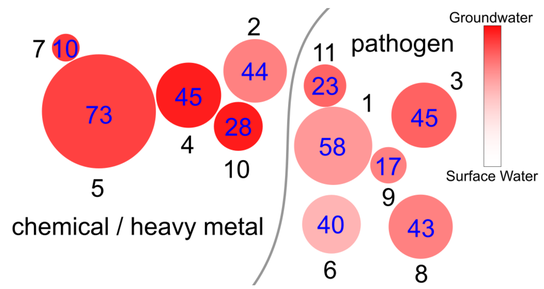

Fig. 3 Water type classification and associated diseases. No. of articles are shown in the cluster circles.

- 98% of the retrieved articles are within the scope of “groundwater”;
- Most chemical related articles study groundwater only;
- Pathogen related articles also include surface water in addition to groundwater.

No. of Article Growth Patterns

- Numbers of articles are increasing over the past 30 years;
- The growth rate is increasing;
- Articles talking about chemicals grew slower before 2013, but their numbers surpassed pathogen related articles in recent years.

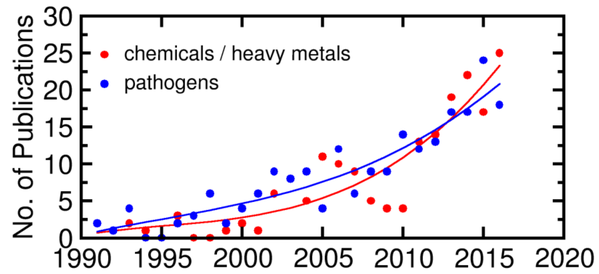

Fig. 4 The growth patterns of the numbers of articles

Conclusions

- Research on water-related diseases in groundwater focuses mainly on chemicals and pathogens;
- Cancer and diarrhea are two major diseases associated with chemical and pathogen studies respectively;
- Research on groundwater diseases is also related to surface water studies;
- Research on groundwater disease is growing.