


May 20th, 12:30 PM

# Automatic Article Screening in Systematic Review

Tong Wang

*University of Massachusetts Boston*

Follow this and additional works at: [http://escholarship.umassmed.edu/cts\\_retreat](http://escholarship.umassmed.edu/cts_retreat)

 Part of the [Bioinformatics Commons](#), [Databases and Information Systems Commons](#), [Library and Information Science Commons](#), [Theory and Algorithms Commons](#), and the [Translational Medical Research Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#).

---

Wang, Tong, "Automatic Article Screening in Systematic Review" (2014). *UMass Center for Clinical and Translational Science Research Retreat*. 105.

[http://escholarship.umassmed.edu/cts\\_retreat/2014/posters/105](http://escholarship.umassmed.edu/cts_retreat/2014/posters/105)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in UMass Center for Clinical and Translational Science Research Retreat by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

## Automatic Article Screening in Systematic Review

Tong Wang

Computer Science Department, UMB

[wintor12@gmail.com](mailto:wintor12@gmail.com)

In recent years, evidence-based medicine has become increasingly important in guiding health care practices. Systematic review, which is the core component of evidence-based medicine, attempts to identify and synthesize all the empirical evidence from online resources such as pubmed to answer a given research question. Usually a clinical researcher needs to choose dozens of related articles as references to work on systematic review. However, there are usually thousand of articles retried from pubmed after keywords searching, it is time consuming to read each of the articles to find the right ones.

My work is to apply text mining and machine learning techniques to screen articles automatically, to minimize the articles set without losing any right ones. This project is processing from three aspects: words, sentences and articles. Words are analyzed by counting term frequency, sentences are by parsing syntactic structures and semantic analyzing, articles are by analyzing general features like author, how many articles reference it. And the data sets are imbalance since the 'right' articles are only a very small part. So there are a lot of challenges to be addressed.

My work is currently focusing on words level, to try different feature selection methods and classifiers to improve the performance.