



## Doctoral Dissertations 2014-current

Off-campus UMass Amherst users: To download campus access dissertations, please use the following link to [log into our proxy server](#) with your UMass Amherst user name and password.

Non-UMass Amherst users: Please talk to your librarian about requesting this dissertation through interlibrary loan.

Dissertations that have an embargo placed on them will not be available to anyone until the embargo expires.

### A Proportionality-based Approach to Search Result Diversification

[Download](#)

[SHARE](#)

[Van Bac Dang, University of Massachusetts - Amherst](#)

[Follow](#)

Date of Award  
Spring 5-9-2014

Document Type  
Open Access Dissertation

Degree Name  
Doctor of Philosophy (PhD)

Degree Program  
Computer Science

First Advisor  
W. Bruce Croft

Keywords  
search result diversification, proportionality, term level diversification, clustering

Subject Categories  
Computer Sciences

Abstract  
Search result diversification addresses the problem of queries with unclear information needs. The aim of using diversification techniques is to find a ranking of documents that covers multiple possible interpretations, aspects, or topics for a given query. By explicitly providing diversity in search results, this approach can increase the likelihood that users will find documents relevant to their specific intent, thereby improving effectiveness. This dissertation introduces a new perspective on diversity: diversity by proportionality. We consider a result list more diverse, with respect to some set of topics related to the query, when the ratio between the number of relevant documents it provides for each of these topics matches more closely with the topic popularity distribution.

Enter search terms:

  

[Advanced Search](#)

[Notify me via email or RSS](#)

[Browse](#)

[Collections](#)

[Disciplines](#)

[Authors](#)

[Author Corner](#)

[Author FAQ](#)

[Submit Dissertation](#)



Consequently, we derive an effectiveness measure based on proportionality and propose a new framework for optimizing proportionality in search results, which we show to be more effective than existing techniques. Diversification would be impractical without the ability to automatically infer the set of topics associated with the user queries. Therefore, we study cluster-based techniques for generating these topics from publicly available data sources. Based on the challenges that we observe with topic generation, we present a simplified term-based representation for query topics. Specifically, we propose to identify for each query a single set of terms that describes its topics. This set is provided to a diversification technique which in effect treats each of the terms as a topic to determine coverage in the search results. We call this approach term level diversification and we show that it can promote diversity with respect to the topics underlying the input terms. This simplifies the task of finding a set of query topics, which has proven difficult, to finding only a set of terms. We also present a technique as well as several data sources for generating these terms effectively.

Recommended Citation

Dang, Van Bac, "A Proportionality-based Approach to Search Result Diversification" (2014). *Doctoral Dissertations 2014-current*. Paper 66.  
[http://scholarworks.umass.edu/dissertations\\_2/66](http://scholarworks.umass.edu/dissertations_2/66)



This page is sponsored by the [University Libraries](#).

© 2009 [University of Massachusetts Amherst](#) • [Site Policies](#)

 **DIGITAL COMMONS\***  
powered by bepress