

Diversity and Novelty in Information Retrieval

Rodrygo L. T. Santos
Univ. Federal de Minas Gerais
Belo Horizonte, MG, Brazil
rodrygo@dcc.ufmg.br

Ismail Sengor Altıngöve
Middle East Technical University
Ankara, Turkey
altıngöve@ceng.metu.edu.tr

Pablo Castells
Univ. Autónoma de Madrid
Madrid, Spain
pablo.castells@uam.es

Fazli Can
Bilkent University
Ankara, Turkey
canf@cs.bilkent.edu.tr

ABSTRACT

This tutorial aims to provide a unifying account of current research on diversity and novelty in different IR domains, namely, in the context of search engines, recommender systems, and data streams.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*search process*

Keywords

Relevance, Diversity, Novelty, Ambiguity, Redundancy

1. OVERVIEW

Information retrieval has been traditionally approached as a pursuit of relevant information, under the assumption that the users' information needs are unambiguously conveyed by their information requests. While such an assumption may have arguably held in the library setting where the early studies of relevance-oriented ranking were conducted, it does not hold in general, and it is unlikely to hold in particular for the multitude of users' needs in modern information retrieval systems, such as search engines and recommender systems. In order to identify relevant information under the uncertainty posed by the users' requests, an effective approach is to diversify the retrieved results. By doing so, an IR system can minimise the chance of wrongly guessing the users' needs, which may cause the users to abandon their retrieval task.

Through a stream of active research and experiences, diversity and novelty can be said to have by now consolidated into a significant body of techniques, methodologies, theories, and knowledge in the field of information retrieval. This tutorial aims to provide a unifying account of current research on diversity and novelty in different IR domains. In

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright is held by the owner/author(s).

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.

ACM 978-1-4503-2034-4/13/07.

particular, we will cover the motivations, as well as the most established approaches for producing and evaluating diverse results in the context of search engines, recommender systems, and data streams. By contrasting the state-of-the-art in these multiple domains, this tutorial aims to derive a common understanding of the diversification problem and the existing solutions, their commonalities and differences, as a means to foster new research directions.

In particular, the tutorial attendees will:

- understand the importance and complexities of achieving diversity/novelty for various IR domains;
- learn the state-of-the-art approaches for diversity/novelty in search results, documents and streaming data, and recommender systems;
- learn the fundamental evaluation metrics and have an overview of past and current evaluation campaigns
- get an overview of other related application areas that include query suggestions, image search, aggregated search, spatial object retrieval;
- obtain a unified view of the topic as a take-away message; i.e., the connections between various methods employed in different domains as well as the differences between them.

2. TUTORIAL OUTLINE

1. Practical and Theoretical Background
2. Diversity in Search
 - *Implicit and Explicit Diversification Approaches*
 - *Diversity Evaluation*
3. Diversity in Recommendation
 - *Problem Statement*
 - *Novelty and Diversity Enhancement*
 - *Novelty and Diversity Evaluation*
4. Diversity in Documents and Streams
 - *Document-level Novelty*
 - *Novelty and Diversification of Document Streams*
 - *Evaluation*
5. Other Application Areas