

Supporting FRBRization of Web Product Descriptions

Naimdjon Takhirov, Fabien Duchateau,
Trond Aalberg

Department of Computer and Information Science
Norwegian University of Science and Technology

Theory and Practice of Digital Libraries (TPDL' 2011)
Berlin, Germany

Outline

- 1 Introduction
 - Motivation
 - Overview of FRBR
- 2 FRBRization of Web Product Descriptions
- 3 Experiments
 - Experiment Protocol
 - Results
 - Issues
- 4 Conclusion

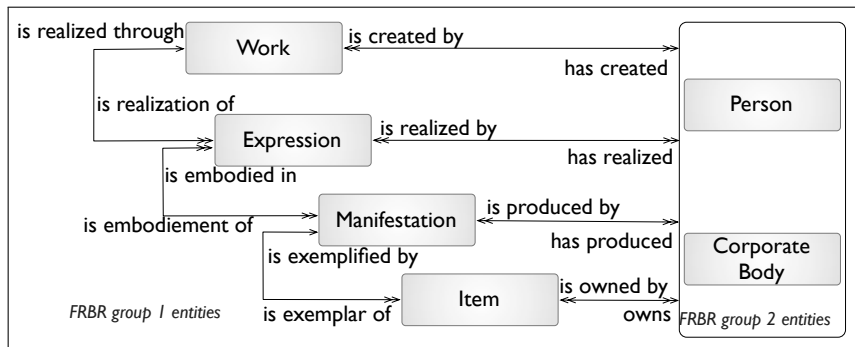
Motivation

- The Web is a major source of information for many users
- Larger amount of information than in library catalogs
- Web product descriptions can be interpreted at manifestation level (as a MARC record describing a publication)
- Enabling discovery of unknown, potentially interesting items
- Supporting more *knowledge-like* representation
- Various sources can be checked for verification

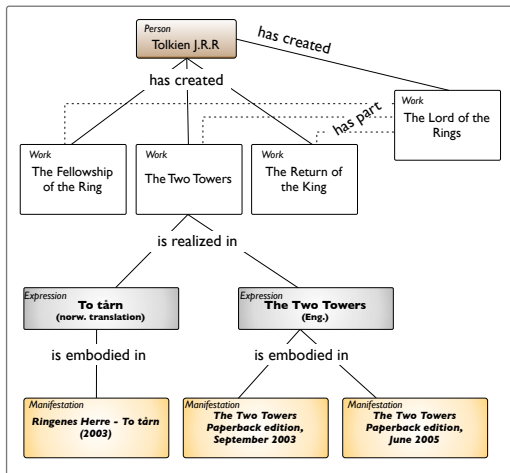
Overview of FRBR

- The goal is to establish a precise and shared understanding of what the bibliographic record should provide information about
- An entity-relationship model
- FRBR is a conceptual model
 - Not a specific metadata schema or data model
 - On the other hand, the conceptual model we use should be the foundation for the the logical data model
- A lot of experiments on using FRBR so far, but no clear agenda for realizing the model in library systems

Overview of FRBR (2)

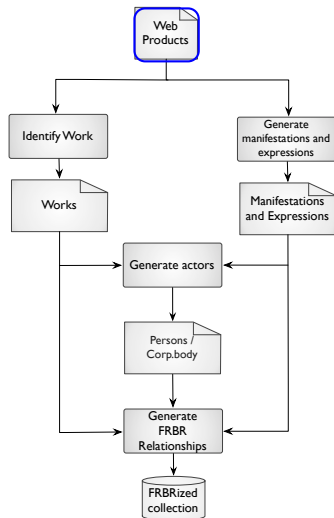


Overview of FRBR (3)



Workflow

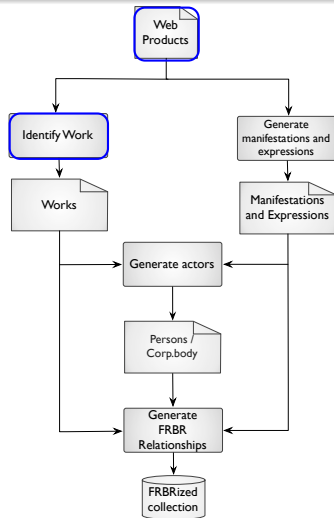
- Set of input product descriptions
- Different properties as compared with bibliographic records
- Interpret at manifestation level



Workflow

Identify work

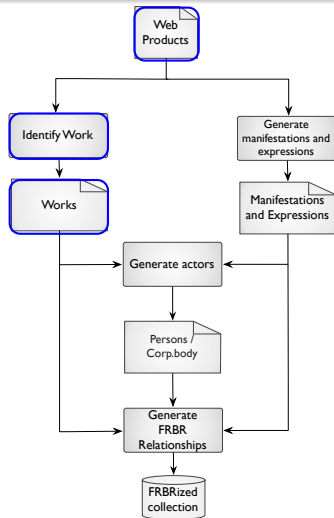
- Local generation method
- Using an external service (e.g., OCLC Classify)
- Retrieve records and apply existing FRBRization techniques



Workflow

Identify work

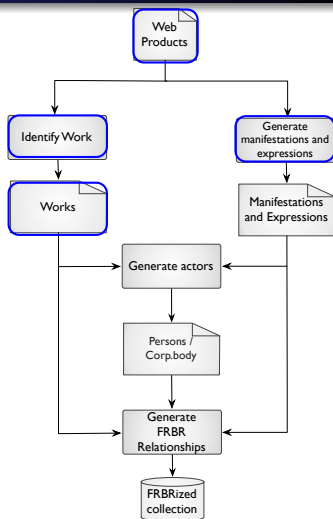
- Local generation method
- Using an external service (e.g., OCLC Classify)
- Retrieve records and apply existing FRBRization techniques



Workflow

Generate manifestations and expressions

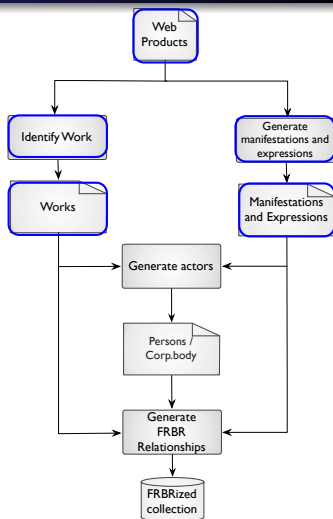
- Obtain all manifestations (e.g. xISBN, ThingISBN)
- Use one of the expression level attributes for clustering/identification (e.g., language)
- Generate local identifiers
- Store references to manifestations



Workflow

Generate manifestations and expressions

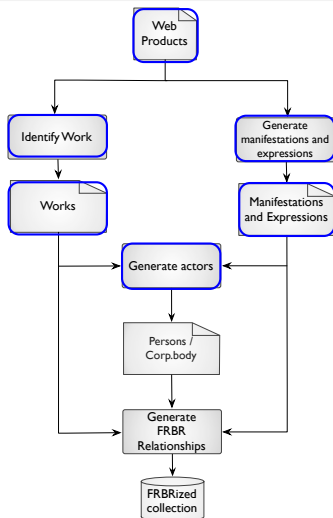
- Obtain all manifestations (e.g. xISBN, ThingISBN)
- Use one of the expression level attributes for clustering/identification (e.g., language)
- Generate local identifiers
- Store references to manifestations



Workflow

Generate actors

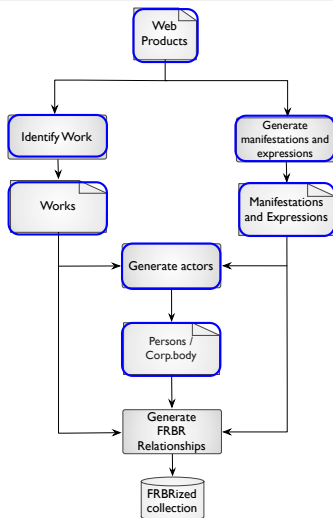
- Actor: responsible for the work, such as author, director
- Use VIAF for search and perform string matching on the results
- Choose the top-1 from the list of top 30 hits (ranking is based on an average of Monge Elkan, Jaro Winkler and Levenshtein similarity values)



Workflow

Generate actors

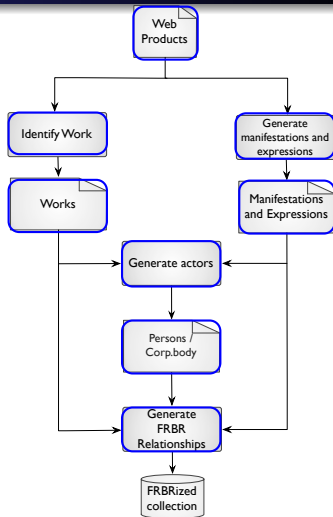
- Actor: responsible for the work, such as author, director
- Use VIAF for search and perform string matching on the results
- Choose the top-1 from the list of top 30 hits (ranking is based on an average of Monge Elkan, Jaro Winkler and Levenshtein similarity values)



Workflow

Generate FRBR relationships

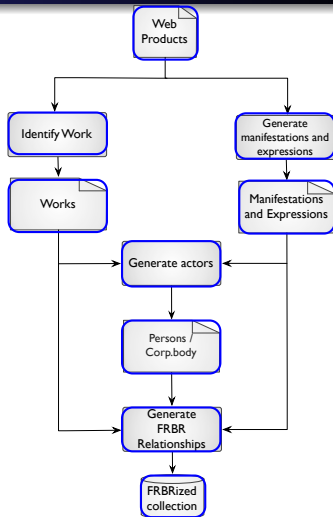
- Use locally generated identifiers
- Link relationships between FRBR entities
- Generate output file (HTML, RDF/XML, XML, SQL insert statements etc)



Workflow

Generate FRBR relationships

- Use locally generated identifiers
- Link relationships between FRBR entities
- Generate output file (HTML, RDF/XML, XML, SQL insert statements etc)



Experimental Protocol

- 80 best selling fiction authors (extracted from Wikipedia)
- Amazon Product Advertising API for collecting dataset (ItemSearch operation on Book, Video, Music indexes)
- Attributes: title, author (director for movies), various contributors, ISBN, language, and release date
- Services: OCLC Classify, xISBN, VIAF

Results

FRBR Entity Type	# Input Resources	# Discov. Entities	# FRBRized Entities
<i>Work</i>		739	684
<i>Expression</i>			5074
<i>Manifestation</i>	1656	28245	28245
- <i>Book</i>	856	27588	27588
- <i>Video</i>	102	542	542
- <i>DVD</i>	190	113	113
- <i>Music</i>	508	2	2
<i>Actor</i>	2221	1569	2221

FRBRPedia

<http://j.mp/frbrpedia>

The screenshot shows the FRBRPedia web application. At the top, there is a navigation bar with links for Home, About, Plugin, FRBR, and DBpedia. Below the navigation bar, there is a search area with the text "ISBN or Amazon ASIN:" followed by an input field containing "0618640150" and a "Submit" button. To the right of the input field, it says "or click [here](#) for an example." Below the search area, there is a row of buttons: "Visual FRBR", "Work.rdf", "Expression.rdf", "Manifestation.rdf", and "DBpedia". The "DBpedia" button is highlighted. Below the buttons, there is a list of results for the ISBN 0618640150:

- Rank: 1 [The Lord of the Rings](#)
- Rank: 2 [The Lord of the Rings: The Two Towers](#)
- Rank: 3 [The Lord of the Rings film trilogy](#)
- Rank: 4 [The Lord of the Rings: The Fellowship of the Ring](#)
- Rank: 5 [The Lord of the Rings: The Return of the King](#)

To the right of the search area, there is a text block that reads: "FRBRPedia is a tool to FRBRize web products. As a case study, we implemented this functionality against Amazon database." Below this, there is a section titled "Installation steps for the plugin:" with a list of three steps:

1. [Install CreaseMonkey](#)
2. [Download frbr2amazon plugin](#)
3. Browse Amazon product display page, e.g.: [here](#)

Below the installation steps, there is a section titled "Contact information" with the following details:

Naimdjon Takhirov
takhirov@idi.ntnu.no

Fabien Duchateau
fabien@idi.ntnu.no

At the bottom right of the page, there is a logo for NTNU – Trondheim, Norwegian University of Science and Technology.

Issues

- Search results from Amazon often needed to be cleaned (e.g., “The Lord of the Rings: The Return of the King (Widescreen Edition)”)
- Missing work information (e.g., the Classify API database did not include requested item)
- Several authors in VIAF (duplication problem)

Conclusion

- FRBR is not limited to library catalogs
- Applied FRBR model beyond its typical realm
- A methodology with different services to FRBRize product descriptions
- Enables easier discovery of potentially interesting items
- Our experiment show good results
- What about more complex cases (e.g. aggregate works)?

Questions or Comments?

Supporting FRBRization of Web Product Descriptions

Naimdjon Takhirov

takhirov@idi.ntnu.no