Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks

Leon Burkard (burkard@kit.edu), Andreas Geyer-Schulz
Outline

- A traditional publication process
- Challenges & Issues
- Use-Case & Requirements for a literature middleware
- Demonstration of the LitObject-Middleware
- Cogeneration of citation networks
- Future Work
- Summary
A traditional publication process

- Editorial sub-processes
- Review sub-processes
- Production editor sub-processes

- Literature work
- Writing / Citing
- Literature Management

- Final typesetting
- Offline publication processes
- Ordering & shipment processes
- DOI Handling, reference & metadata reparation
- Classification, keywords, indexing

- Retrieval / Indexing
- Download in personal literature database.
- Import of document’s reference metadata (BibTeX, RIS, etc.)
Challenges

- How to **rediscover** past literature searches?
- How to **share** and **persist** literature searches?

**Scientific Quality Management**

- Time-consuming **reference and content checks**
- How to **store cited literature** permanently?

**Distribution**

- How to provide “good” **reference metadata and links to attachments**?

**Consumption / Analysis**

- How to collect and use collected metadata for further usage e.g. **citation networks**?
- How to consume “good” **reference metadata and links to attachments** with various literature management approaches?
a) Manual typing of reference, manual linking/importing attachments

b) Attempt to extract information out of the document.
   → Combination with a google search.

c) Automatic import
   a) Crawler / Website Translators
   b) Embedded metadata in the website
Zotero [zoh-TAIR-oh] is a free, easy-to-use tool to help you collect, organize, cite, and share your research sources.

Grab your research with a single click.

A personal research assistant.

Zotero is the only research tool that automatically senses content in your web browser, allowing you to add it to your personal library with a single click. Whether you’re searching for a preprint on
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks
Literature can be organized in collections

Attached document

Imported reference
Multiple downloads possible in one step
Multiple downloads possible in one step

![Multiple downloads possible in one step](image-url)
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks

08.07.2014
Various export options

Quick Export:
1. Select one or many literature entries in Zotero.
2. Press Ctrl+Shift+C for copying the reference to the clipboard.
3. At the desired destination document press Ctrl+V (paste) and the corresponding literature entries’ references are exported in the selected style.
Example BibTeX export

@inproceedings{geyer-schulz_apl-tutoring_1989,
  address = {New York, {NY}, {USA}},
  series = {{(APL) '89}},
  title = {An (APL)-tutoring Adventure Game},
  isbn = {0-89791-327-2},
  url = {http://doi.acm.org/10.1145/75144.75165},
  doi = {10.1145/75144.75165},
  abstract = {This paper is concerned with the development of a tool to embed the learning component in an object oriented adventure game. The APL-CAE, is is a freely definable, using a game model approach with a number of exercises that are entered into the system and are analysed by the teachers, who use it to improve the learning process. This book is part of the proceedings of the 1989 APL conference. 1989, pages = {148-157},
}

@article{franke_update_2009,
  title = {An update algorithm for restricted random walk clustering},
  volume = {3},
  issn = {1862-5347, 1862-5355},
  doi = {10.1007/s11634-009-0039-6},
  abstract = {In this article, we present a randomized dynamic clustering algorithm that given good results in past studies. We discuss different approaches for the problem and present an efficient algorithm for a small percentage of changes in the data set and thus no need for any reclustering ever. In addition, the method is easily parallelizable.},
  number = {1},
  urldate = {2014-06-25},
  journal = {Advances in Data Analysis and Classification},
  author = {Franke, Markus and Geyer-Schulz, Andreas},
  month = jun,
  year = {2009},
}

@incollection{geyer-schulz_social_2012,
  series = {{Communications in Computer and Information Science}},
}

But....what about collaborative working with literature?

08.07.2014
Use-Case Scenario

- Rediscover
- Share
- Persist
- Less duplication of work

- Continuous transfer of literature objects
- Basis for growing, co-generated linked literature objects

Local literature management software

- Annotate
- Ordering / Organizing
- Reference- and Attachment management

LitObject-Middleware

- Import
- Export
- Display entries

Literature Database (organization wide)

- Permanent storage for documents
- Advanced index & retrieval functions

Export

Author 1

- Search
- Store
- Organize
- Export

Author 2

uses

Import

Local literature management software

Author 1

uses

Local literature management software
Requirements for the LitObject Middleware

- **Save literature objects**
  - Reference in a data format, e.g. BibTeX
  - One or many attachments
  - Notes & Annotations to the literature object

- **Save links** between literature objects
  - Literature object a references b,c,d

- **Save collections** of literature objects
  - All literature to a sub-topic

![Diagram showing references and collections]
Requirements for the LitObject Middleware

- Save **literature objects**
- Save **links** between literature objects
- Save **collections** of literature objects
Why not use one of the integrated solutions?

- Commercial services = Limited research possibilities
  - Closed source, difficult to extend for new applications
  - Often too expensive for intensive collaborative literature work
- Vendor Lock-In
  - No literature object sharing, interoperability nearly non-existent
- Extensibility very limited
  - No further services like citation graph generation or recommendation services
  - No seamless integration in publication processes
- We need a solid basis for further development projects in the field of open library and information services
Demo Part 2
- Save literature objects
- Save links between literature objects
- Save collections of literature objects

LitObject Middleware
- Developed as a Mozilla Firefox extension
- Extends the literature management software Zotero
  - Provides the export process step to the LitObject Middleware Service
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks

LitObject Preferences

Please enter the URL to the LitObject middleware service and a username, e.g. your student account. At current development status no password is required.

LitObject Server Preferences

Server: http://localhost:9000
Username: test1

Synchronisation Preferences

Please select all literature collections that should be pushed to the LitObject service:

- [x] demo1
- [x] demo2
- [x] demo3

[Cancel] [Save]
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks
Hi demo1

Select one of your exported collections:

- demo1
- demo3
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks

- **year**: 2009
- **urldate**: 2014-06-25
- **title**: An update algorithm for restricted random walk clustering for dynamic data sets
- **pages**: 63–92

- **incollection**: A Social Location-Based Emergency Service to Eliminate the Bystander Effect
- **publisher**: Springer Berlin Heidelberg
- **isbn**: 978-3-642-25205-1, 978-3-642-25206-8
- **language**: en
- **copyright**: ©2012 Springer-Verlag (GmbH) Berlin Heidelberg
- **author**: Geyer-Schulz, Andreas and Ovelgrünne, Michael and Sonnenbichler, Andreas C.
- **url**: http://link.springer.com/chapter/10.1007/978-3-642-25206-8_7
- **series**: Communications in Computer and Information Science
- **abstract**: The availability of inexpensive smartphones with (GPS) units and the ability to run 3rd party software facilitated the rapid emergence of location-based services (LBS) related application scenarios, in this contribution we want to motivate the use of mobile devices for personal safety services. The obstacles in the social help process have been summarized in the concept of the bystander effect. This contribution aims to show how a (LBS) can service that notifies nearby contacts of a victim of the incident and guides them to the victim so that they can provide help. Furthermore, we will discuss the benefits of participating from different sources of data. Also, we will discuss and compare approaches for the system design we consider the benefits in practice.
- **title**: A Social Location-Based Emergency Service to Eliminate the Bystander Effect
- **number**: 222
- **pages**: 112–130
- **keywords**: Algorithm Analysis and Problem Complexity, Bystander effect, Coding and Information Theory, Computer Communication Networks, Computer Games, multimedia, Information Systems, Privacy, Social network analysis
- **editor**: Obaidat, Mohammad S. and Tsihrintzis, George A. and Filipe, Joaquim
- **year**: 2012
- **urldate**: 2014-06-25
- **month**: #jan#
- **booktitle**: e-Business and Telecommunications

- **inproceedings**: An (APL)-tutoring Adventure Game
- **publisher**: (ACM)
- **doi**: 10.1145/75144.75165
- **isbn**: 0-89791-327-2
- **author**: Geyer-Schulz, A. and Taudes, A. and Mitlöhner, J.
- **url**: http://doi.acm.org/10.1145/75144.75165
- **series**: (APL) ’89
- **abstract**: This paper is concerned with the development of a program system that assists the teaching of (APL)2. Rather than presenting the teaching oriented adventure game environment written itself in {APL}2. The game's actors, objects and places, which are implemented as hierarchically classified by a game simulator. The teaching is integrated as follows: The student/teacher is a specific actor who is presented with a number of established questions. The solutions entered into the system are analysed by the teachers, who use an (LR)(1)-parser generator and "critic", which help the student to judge the quality of its actions.
- **title**: An (APL)-tutoring Adventure Game
- **year**: 1989
- **pages**: 148–157
- **file**: content/home/leon/mozilla/firefox/atcfxfq7.Demo_LIS Workshop/zotero/storage/NB5DZ666/content.pdf:application/pdf
- **address**: New York, (NY), (USA)
- **urldate**: 2014-06-25
- **booktitle**: Conference Proceedings on (APL) As a Tool of Thought

- **article**: (GfK)-Konsumklima: Deutsche in Kaufstimmung, Sparen ist out
- **shorttitle**: (GfK)-Konsumklima
- **title**: (GfK)-Konsumklima: Deutsche in Kaufstimmung, Sparen ist out
Leon Burkard, Andreas Geyer-Schulz - Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks
Challenges

- How to **rediscover** past literature searches?
- How to **share** and **persist** literature searches?
- How to **work together** with regard to literature?
- Do the work of the author(s) again for reviews?
- Time-consuming **reference and content checks**
- What about **interoperability** between various software tools?
- How to **store cited literature** permanently?

**Zotero** A
- ZoteroPlugin
- Import
- Export

**Zotero** B
- ZoteroPlugin

**Mendeley** C
- MendeleyPlugin

**Webservice/API**
- Access data

**LiteratureDB**
- Search & Administration

**Website**
- Import
Challenges

- How to provide “good” reference metadata and links to attachments?
- How to collect and use collected metadata for further usage e.g. citation networks?
- How to consume “good” reference metadata and links to attachments with various literature management approaches?
Cogeneration of citation networks – Generic process

**Data collection**
Zotero

**Data transfer**
LitObject-Middleware

**Data import**
LiteratureDB, Relational Database

**Data analysis & visualization**

**Authors/Users:** Search, retrieve, check and organize literature
→ **Quality management shifted to the individual**

**Client-side software:** Export references and attachments to a web-service for further use
→ **Software supported transfer of structured data (links)**

**Server-side software:** Import to link tables as data basis for citation graph analysis and algorithms
→ **Using the LitObject Middleware for a growing data basis with individuals’ support in checking and organizing**

Using a co-generated growing data basis for analysis, visualization
→ **Not covered today**
Cogeneration of citation networks

Further usage of the permanent datastore

Export

Webservice / API

Access data

Import

Website

LiteratureDB

Search & Administration

Imported article

Cited article

Cited article

Cited article

Imported article

Cited article

Cited article

Cited article
Step-by-Step co-generation

1. Organize literature
   - Save all cited literature in one collection

2. Writing article
   - Save all collected literature for the topic in another collection
Step-by-Step co-generation

3. - Export both collections to the webservice

- Save links between documents
- During the import in the LiteratureDB save additionally a new document - that is the article newly written.
- Save additionally a link from the newly written document (the additional document) to each document in both collections

4. - Export both collections to the webservice
Step-by-Step co-generation

5. Use the growing, cogenerated link database for further applications and visualizations

6. Work in progress:
   - Provide an API for automatic graph generation
   - Provide an API for automatic citation data export in various formats
Future Work

- Imported references are not validated
  - E.g. how to detect/solve misspelled author names?
  - Sometimes reference and attachment do not match

- Duplicates detection is future work
  - Currently limited to a single user duplicates detection on client-side

- User Management exists only on an elementary level

- Process improvements using the LitObject middleware
  - Use the LitObject middleware for advanced publication and review processes.
  - Use the LitObject middleware for private institutional repositories
  - Provide an automated citation network graph generation and structured linked data export.
Future Work

- Support more literature management tools
  - Mendeley is on top of the agenda.
  - An HTML5 online interface for organizing literature is planned.

- Developing new services on top of the LitObject middleware
  - Implement classification algorithms that use the LitObject middleware objects (reference and fulltext document)
  - Implement and test recommendation algorithms.

- Discuss/implement how to provide references and links to fulltext documents for common literature management software

- Publish a small open-source literature database as organization internal persistence unit.
Summary

- The LitObject middleware helps to share literature objects consisting of references and attachments
  - Basis tool for developing advanced processes that include literature
  - Basis tool for developing new applications that use the structured literature data in the webservice

- Possibility to shift workload to authors/literature organizers and reduce duplicate work
  - No manual reference transfer throughout the publication process chain.
  - Chance to build up a database of checked, correct references that can be used from authors again.

- Import step for literature management software through websites needs further development instead of new web translators every time.
  - Maybe unAPI v2
Backup
How to import LitObjects

There exist various metadata formats for exposing metadata to literature management tools through websites:

- CoIns (ContextObjects in Spans)
- DublinCore Metadata tags: dc.*
- Eprints tags: eprints.*, PRISM tags: prism.*, etc.
- MODS XML, Marc21, MAB, etc.
- BibTex, RIS

- Highwire Press Keys: citation_*
  - Field for linking to the fulltext attachment: citation_pdf_url

- Lesson learnt: Either good metadata and no link to literature attachments or bad metadata and links to literature attachments

Good = Export a BibTeX reference to another format, import and convert it back to BibTeX without too much information loss.
A small extension for unAPI

- unAPI by Daniel Chudnow, 2006
  - Link to common reference formats instead of translating to other microformats before

1. Embed a resolver service

```html
<html lang="en">
  <head>
    [...]<link rel="unapi-server" type="application/xml" title="unAPI" href="http://localhost:5050/unapi/demo1" />
  </head>
  [...]</html>
```

2. Link by using identifiers to the service

```html
<abbr class="unapi-id" title="UP9MJ5F6">
  <a href="http://localhost:9000/demo1/.../content_one"></a>
  <a href="http://localhost:9000/demo1/.../content_two"></a>
</abbr>
```

**New:** Add links to documents/fulltext attachments
Zotero’s priority order for choosing importers

<table>
<thead>
<tr>
<th>Method</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Metadata</td>
<td>400</td>
</tr>
<tr>
<td>DOI</td>
<td>300</td>
</tr>
<tr>
<td>ColIns</td>
<td>250</td>
</tr>
<tr>
<td>unAPI / Highwire</td>
<td>200</td>
</tr>
</tbody>
</table>

Specific Web Translator (individual simple crawler for each specific site)

- Embedded Metadata (<meta> keys in <head> section)
- Resolve DOI
- ColIns
- unAPI
Example for ContextObjects in spans

unAPI Details

Get the available formats

GET http://localhost:5050/unapi/demo1?id=7QZM2MXJ

Get the metadata in one of the available formats

http://localhost:5050/unapi/demo1?id=7QZM2MXJ&format=bibtex
A full graph of our private link database