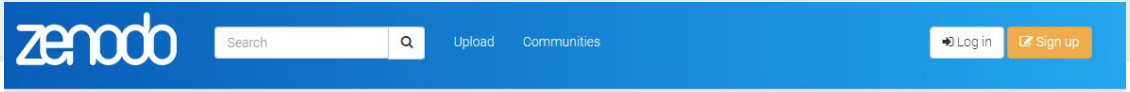
Open Science Platforms as Data Repositories for Automated Summarization

Sarah Frank, CERN and Graz University of Technology
Andreas Wagner, CERN
Christian Gütl, Graz University of Technology

Introduction

- Open science platforms provide a wealth of information
  - Many topics
  - Many forms of content
- Search for relevant papers can be time consuming
- Keeping up with new content can be overwhelming
- Usage should be low-barrier



All versions

Found 64902 results.

< 1 2 3 4 5 6 7 8 9 >

Sort by: Most recent asc.

Access Right

- [x] Open (2627160)
- [ ] Closed (54872)
- [ ] Restricted (9694)
- [ ] Embargoed (1220)

File Type

- [x] Pdf (1139469)
- [ ] Html (425554)
- [ ] Png (404947)
- [ ] Jpg (378336)
- [ ] Zip (136885)
- [ ] Xlsx (33304)
- [ ] Txt (30105)
- [ ] Docx (25876)
- [ ] Csv (24978)
- [ ] Xml (19074)

Keywords

- [ ] Taxonomy (1144922)
- [ ] Biodiversity (1143912)
- [ ] Animalia (777274)
- [ ] Arthropoda (601339)
- [ ] Insecta (420944)
- [ ] Terrestrial (270754)

October 11, 2022 (v1) | Conference paper | Open Access

View

THE IMPORTANCE OF TESTING LABORATORIES AND THEIR ACCREDITATION.

BADALOV UTKIRBEK NOMOZ OGLI;

Standardization is the establishment of a normative unit for product types, models, parameters, dimensions and quality, as well as production technology, testing and control methods, product placement, design, storage, etc. A standard is a normative-technical document on standardization approved by

Uploaded on October 11, 2022

October 11, 2022 (v1) | Conference paper | Open Access

View

HOW TO TEACH ENGLISH LANGUAGE MEDICAL ENGINEERING SPECIALTY STUDENTS

Tukhtasinova Durdona Tolibjonovna;

Students article in medicine d engineering sciences. This manuscript presents the curriculum of the bachelor' at machine-building institute, the spiral approach. on revisiting of topics, with encounter, levels of are increased, new learning to previous learning, and students' is enhanced. A

Uploaded on October 11, 2022

October 11, 2022 (v1) | Conference paper | Open Access

View

THE SIGNIFICANCE OF CERTIFICATION IN ENTERPRISES WITH A QUALITY MANAGEMENT SYSTEM

BADALOV NOMOZ JO'RABOYEVICH;

Involvement of quality system certification experts, registered certification staff and technical experts in the work on quality management system certification. Based on the customer's request to determine the scope of the quality management system and the audit results, the certification offic

Uploaded on October 11, 2022

May 15, 2022 (pdf) | Conference paper | Open Access

View

IJODIY ISHLAR KO'RGAZMASI

FERUZAXON;

TANLOV

Uploaded on October 11, 2022

Keywords

- [ ] Taxonomy (1144922)
- [ ] Biodiversity (1143912)
- [ ] Animalia (777274)
- [ ] Arthropoda (601339)
- [ ] Insecta (420944)
- [ ] Terrestrial (270754)
- [ ] Herbarium (270657)
- [ ] Coleoptera (117930)
- [ ] Chordata (96265)
- [ ] Arachnida (93624)

Key Questions

How to use automated summarization to improve usage and user experience of/on open science platforms?

Automated Summarization - Why?

- Time-saving for the pre-selection phase
- Cuts down on redundancy
- Highlight similarities and differences between papers
- Quick overview over research trends

Automatic Summarization

- Transformer-based models dramatically improved previous results
  - Seems to be the de facto standard in many cases now
- Much focus on news articles in research
  - Most common datasets are made up of them
  - Differ from scientific articles both in length and complexity
- Single-document summarization vs. multi-document summarization

Main requirements

- Selection of relevant text sources
- Extraction of key information
- Summarization for a specified distribution technology
- Personalisation through user specification

Concept I - Summaries as a newsletter

- An easier way of "keeping up to date"

- User specification:
  - Topics of interest through keywords, fields of science, etc.
  - Interval between newsletters
  - Length of summary
  - Communication channel

Concept I - Summaries as a newsletter

- Automatically select papers published
  - in the specified duration of time
  - that fit the chosen topic/field

- Summarize them, factoring in
  - the specified summary length
  - the intended communication channel

Concept II - Summaries a la carte

- If not only the most recent publications are of interest
- Select papers in a tool such as Collaboration spotting X[1]
- Summarize them according to number of selected papers
- Output shown after summarization

[1] Bobic, A., Le Goff, J. M., & Gütl, C. (2021). Collaboration Spotting X-A Visual Network Exploration Tool. In in Proceedings of the The Eighth International Conference on Social Networks Analysis, Management and Security: SNAMS 2021.

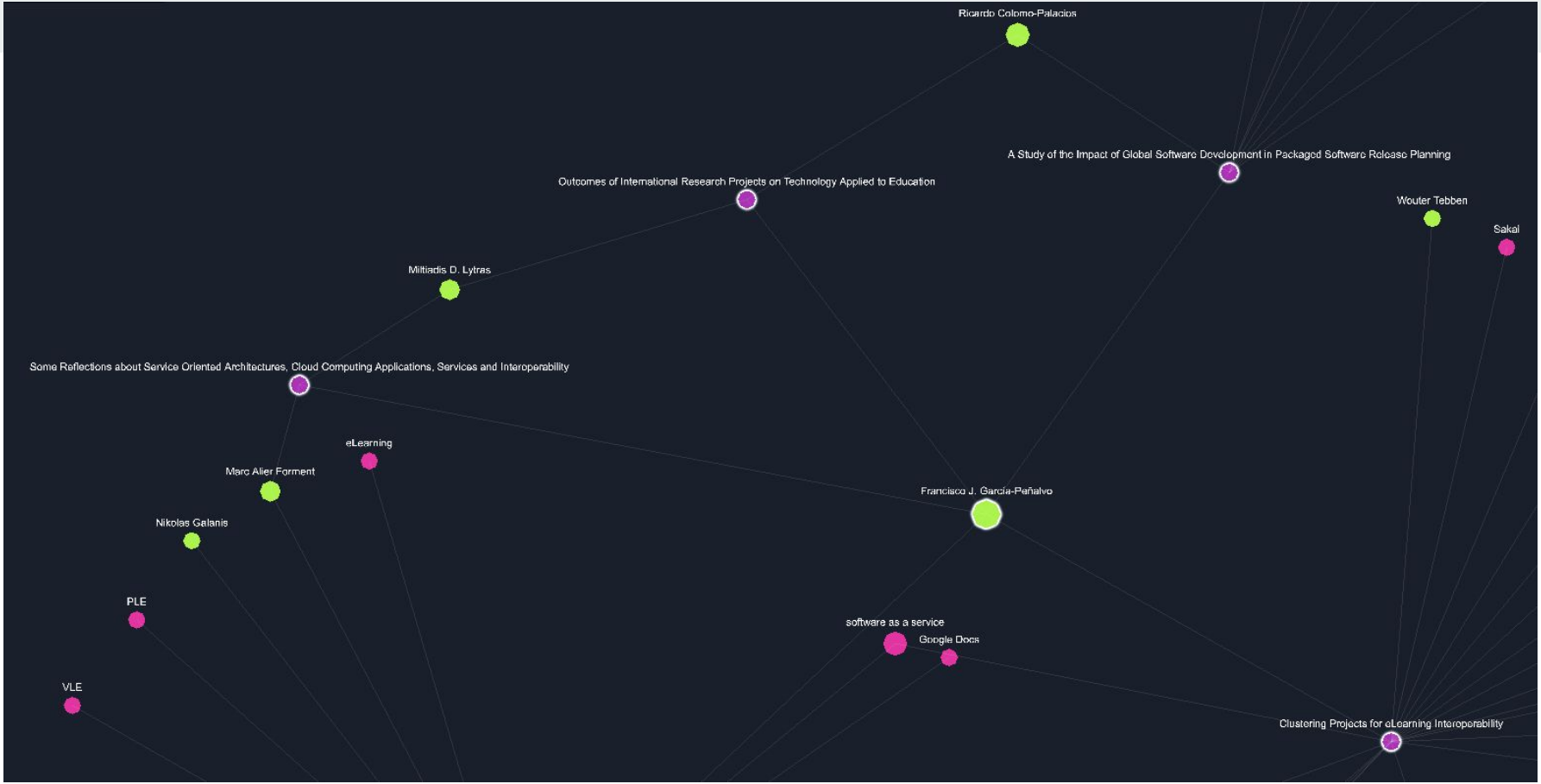


Figure 1. Bobic, A., Le Goff, J. M., & Gütl, C. (2021). Collaboration Spotting X-A Visual Network Exploration Tool. In in Proceedings of the The Eighth International Conference on Social Networks Analysis, Management and Security: SNAMS 2021.

Challenges - User Specification

- Granularity of topic selection
- Availability of topic information
- Given maximum length may render summary unusable

Challenges - Summarization

- Few and/or small datasets for summarization of multiple documents
  - Particularly so for scientific texts, such as papers
- Papers in some domains may be harder to summarize than others
- Staying true to the information in the paper - no "hallucination"

Conclusion

- Keeping up to date with research can be difficult and time consuming
- Regular summaries regarding recent publications are an attempt to reduce information overload
- Automatic summarization could help but presents challenges