Scholarly Navigation on an Open Science Platform

Mohsine Aabid, Simon Dumas Primbault, Patrice Bellot

Scan me

Scholarly Navigation on an Open Science Platform

A Computational Study of OpenEdition's Server Logs

Mohsine Aabid
mohsine.aabid@openedition.org
OpenEditionLab
Laboratoire d'informatique et systèmes de Marseille

Simon Dumas Primbault
simon.dumas-primbault@openedition.org
OpenEditionLab

Patrice Bellot
patrice.bellot@lis-lab.fr
Laboratoire d'informatique et systèmes de Marseille

Background

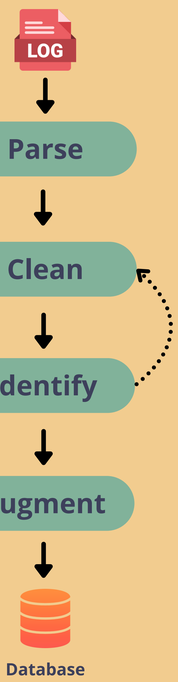
This study analyzes **user behavior** on OpenEdition, a French open access platform for the humanities and social sciences, which includes four services: Journals, Books, Hypothèses (blogs), and Calenda (events), plus a search engine. Using **server logs** from October 2023 and 2024, the research investigates how users interact with and navigate these platforms.

Question

What **patterns** of user behavior are typically observed during navigation within an open access **digital library**?

Data Preprocessing

- Unstructured, noisy data.
- Extract information from log files.
- Clean the non-human queries.
- Identify queries within the same session.
- Enrich with some additional data.
- Structured, enriched data.



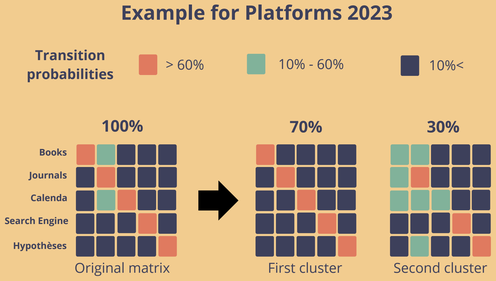
Methodology

- Log files were **preprocessed** to extract features and structure data around identified **user sessions**, enabling a coherent view of user interactions over time.
- **Three separate behavioral analyses** were conducted, each using a different clustering algorithm tailored to a specific objective.

Transitions

- How do users **transition** in the digital library websites?
- Compute **transition matrices** at platform and website levels.
- Cluster sessions via **Expectation Maximization** and derive matrices with cluster proportions.

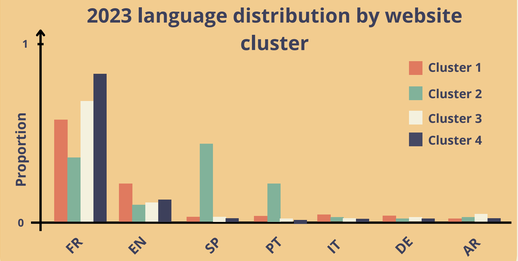
Example for Platforms 2023



Preferences

- Are there some preferences for consulted ressources?
- Build an **item-item** similarity matrix for editor websites based on sessions.
- Cluster similar **websites** and analyze them by **attributes** like language or discipline.

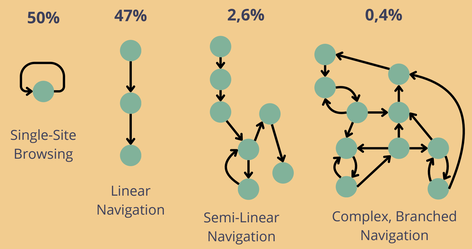
2023 language distribution by website cluster



Topology

- What types of navigation **profiles** can be expected in user sessions?
- Represent visited websites numerically using **Word2Vec**.
- Analyze **topology** of navigation paths and cluster by structure.

Example of navigation strategies



Preliminary Results

- Most sessions stay on one platform; cross-platform sessions often lead to journal sites.
- User preferences emerge in editor metadata—for example, one group favors Spanish and Portuguese sites, another prefers disciplines like sociology, anthropology, education, and area studies
- Navigation patterns reveal four session types: single-site (one editor), linear (no revisits), semi-linear (linear start with later loops), and complex (branched with constant backtracking).

Digital Humanities Research Conference (DH2025)
Sixth Annual Conference, 13–18 July 2025, Lisbon, Portugal
This work has received government funding managed by the Agence Nationale de la Recherche under the Programme France 2030, reference ANR-21-ESRE- 0045