Student Project

# Open Web Index

Hochschule der Medien Stuttgart

Summer Semester 2024

Cem Ardic, Linus Breitenberger, Felice Douglas, Nils Firschau, Sonja Gröschel, Susanne Krol

# Research Questions

What can be learned from Information Science in the development and maintenance of an Open Web Index?

What aspects can be managed similarly to traditional libraries, and what cannot?

Giphy

# Methods of Information Science

- **Cataloging Publications**: Descriptively and by Subject

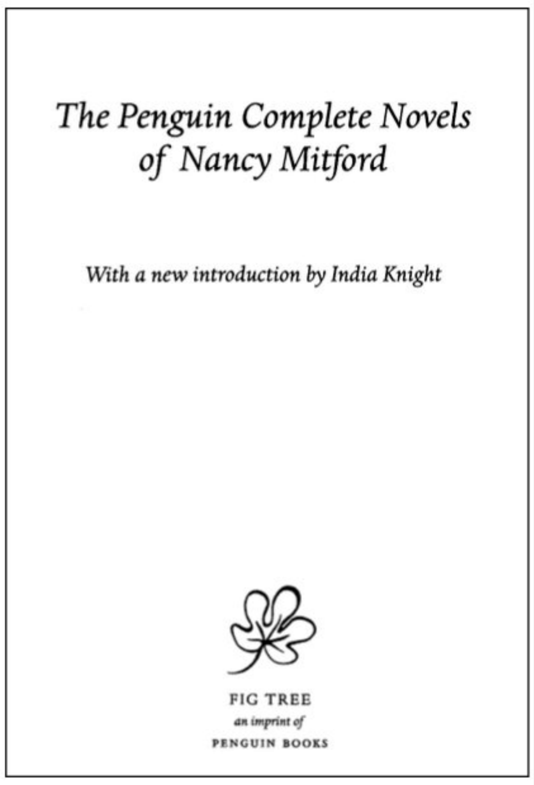
- Subject Headings
- Integrated Authority Files
- Classification
- Metadata

These methods are carried out following **established guidelines** and **standards**.

# Cataloging books

Ohio University Libraries - http://media.library.ohiou.edu

- Manual process
- Guidelines: RDA
- Structured and stable information about the ressource



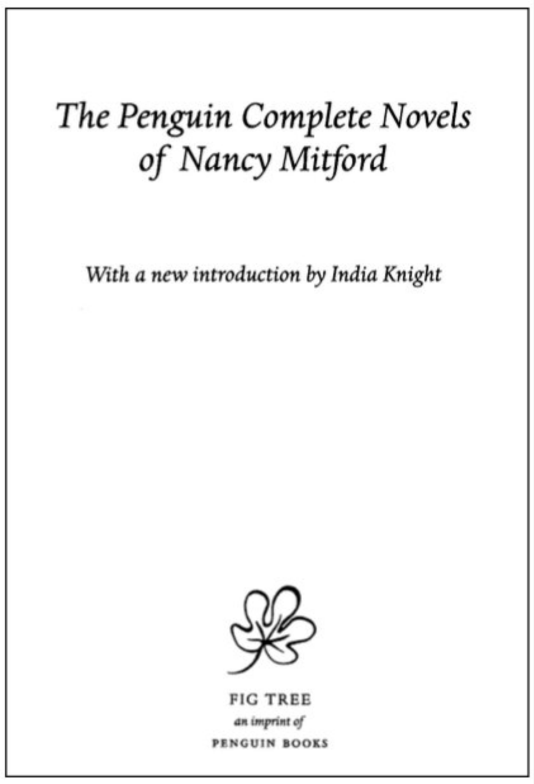
*The Penguin Complete Novels of Nancy Mitford*

*With a new introduction by India Knight*

FIG TREE
*an imprint of*
PENGUIN BOOKS

# Descriptive Cataloging

| | |
|---|---|
| 0500 | Aau |
| 0501 | Text$btxt |
| 0502 | ohne Hilfsmittel zu benutzen$bn |
| 0503 | Band$bnc |
| 1100 | 2011 |
| 1500 | eng |
| 1505 | $erda |
| 2000 | 978-1-905-49089-9$fFesteinband |
| 2000 | 978-1-905-49090-5$fBroschur |
| 3000 | !PPN!Mitford, Nancy*1904-1973*$BVerfasserIn$4aut |
| 3210 | Romane |
| 4000 | The @Penguin complete novels of Nancy Mitford $hwith a new introduction by India Knight |
| 4030 | London, England$nFig Tree |
| 4060 | ix, 975 Seiten |
| 4062 | 24 cm |



*The Penguin Complete Novels of Nancy Mitford*

*With a new introduction by India Knight*

FIG TREE
*an imprint of*
PENGUIN BOOKS

# Descriptive Cataloging

| | |
|---|---|
| 0500 | **Aau** |
| 0501 | Text$btxt |
| 0502 | ohne Hilfsmittel zu benutzen$bn |
| 0503 | Band$bnc |
| 1100 | **2011** |
| 1500 | **eng** |
| 1505 | $erda |
| 2000 | 978-1-905-49089-9$fFesteinband |
| 2000 | 978-1-905-49090-5$fBroschur |
| 3000 | !PPN!**Mitford, Nancy***1904-1973*$BVerfasserIn$4aut |
| 3210 | Romane |
| 4000 | The @Penguin complete novels of Nancy Mitford $hwith a new introduction by India Knight |
| 4030 | **London, England**$nFig Tree |
| 4060 | ix, 975 Seiten |
| 4062 | 24 cm |

What role could **mandatory core metadata elements** play in web documents stored within an open web index, and how might they influence discoverability and interoperability?

Which **core elements** are essential for effectively describing and indexing web documents in an open web index and how can they be gathered?

# Cataloging by Subject

- Structures collections thematically
- Guidelines: RSWK
- Categories:

  **f** Form  
  **g** Geographical  
  **k** Corporate body  
  **p** Personal name  
  **s** Subject  
  **t** Title  
  **v** Conference  
  **z** Time-related

# Cataloging by Subject

- Authority Files: **GND**
- Standardized terminology ensure accuracy and consistency and allows Linked Data

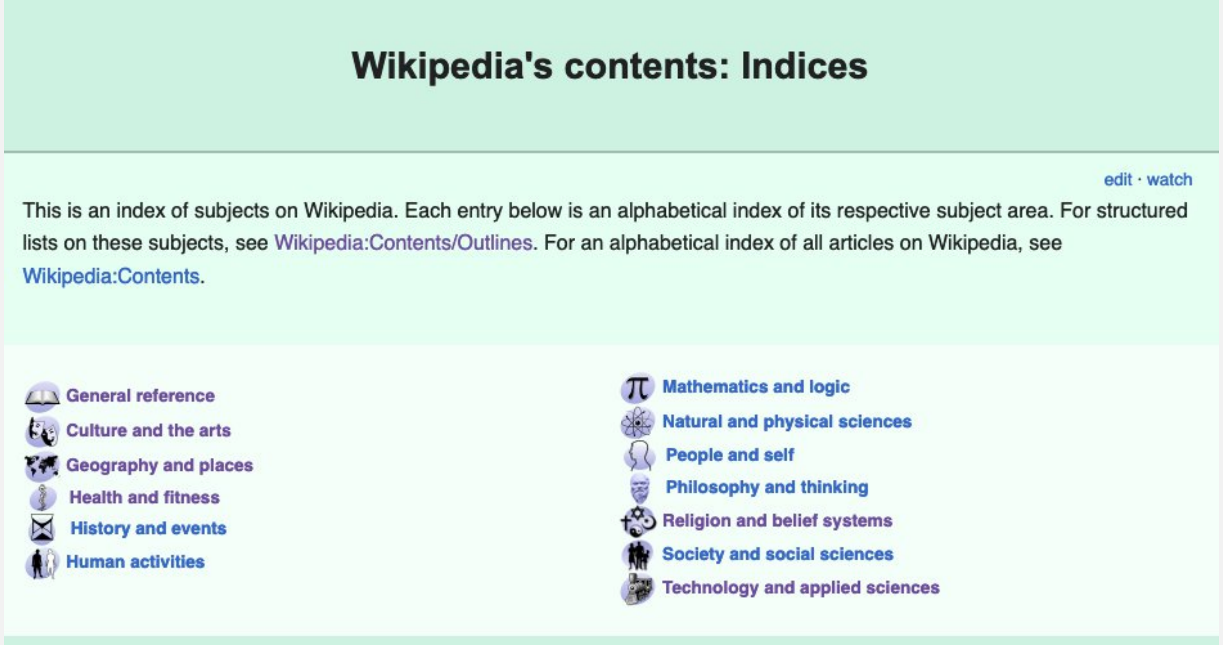
- Subject Headings: **LCSH**

# Classification

- Standardized rules and systems & controlled vocabulary
- Mostly hierarchical with main classes and subclasses

- Classification Systems:
    - Universal Decimal Classification
    - Dewey Decimal Classification
    - Library of Congress Classification
    - Regensburger Verbundklassifikation

# Thematic content organization



## Wikipedia's contents: Indices

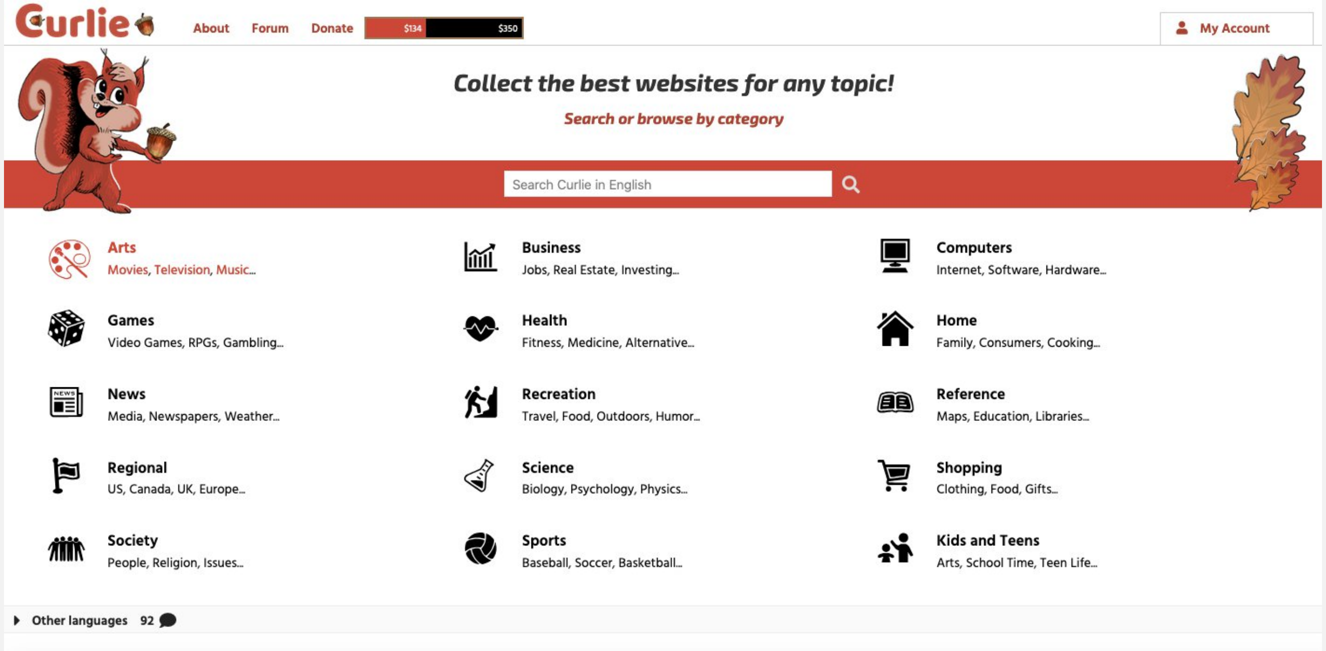
edit · watch

This is an index of subjects on Wikipedia. Each entry below is an alphabetical index of its respective subject area. For structured lists on these subjects, see Wikipedia:Contents/Outlines. For an alphabetical index of all articles on Wikipedia, see Wikipedia:Contents.

- General reference
- Culture and the arts
- Geography and places
- Health and fitness
- History and events
- Human activities

- Mathematics and logic
- Natural and physical sciences
- People and self
- Philosophy and thinking
- Religion and belief systems
- Society and social sciences
- Technology and applied sciences

# Thematic content organization



**Collect the best websites for any topic!**

*Search or browse by category*

**Arts**
Movies, Television, Music...

**Business**
Jobs, Real Estate, Investing...

**Computers**
Internet, Software, Hardware...

**Games**
Video Games, RPGs, Gambling...

**Health**
Fitness, Medicine, Alternative...

**Home**
Family, Consumers, Cooking...

**News**
Media, Newspapers, Weather...

**Recreation**
Travel, Food, Outdoors, Humor...

**Reference**
Maps, Education, Libraries...

**Regional**
US, Canada, UK, Europe...

**Science**
Biology, Psychology, Physics...

**Shopping**
Clothing, Food, Gifts...

**Society**
People, Religion, Issues...

**Sports**
Baseball, Soccer, Basketball...

**Kids and Teens**
Arts, School Time, Teen Life...

Do the thematic structures used by Curlie and Wikipedia provide the best approach for organizing an open web index, or could **other taxonomies** offer greater effectiveness for search and representation of the ?

# Methods of Information Science

Manual processes **not fully applicable** to an Open Web Index and its data due to its scale and dynamic nature.

**→ Scalability of the processes related to content organization and thematic structuring in an Open Web Index?!**

How can **topic modeling** be used for cataloging and content indexing?

What advantages do **vector databases** offer for managing large-scale data in the context of an open web index?

# Website for the Rating System

# Subpages of Rating Website

### Dashboard

Dashboard page with widgets and statistics

### User Profile

Personal Info,
Level & Experience Points,
Badges (Achievements),
Completed Courses

### Rating-Page

Rating mode with rating panel,
Overview of recent ratings (up/down-votes)

### Courses & Quests

To get to the next level, you have to complete quests (tasks).
Completing a course is a quest.

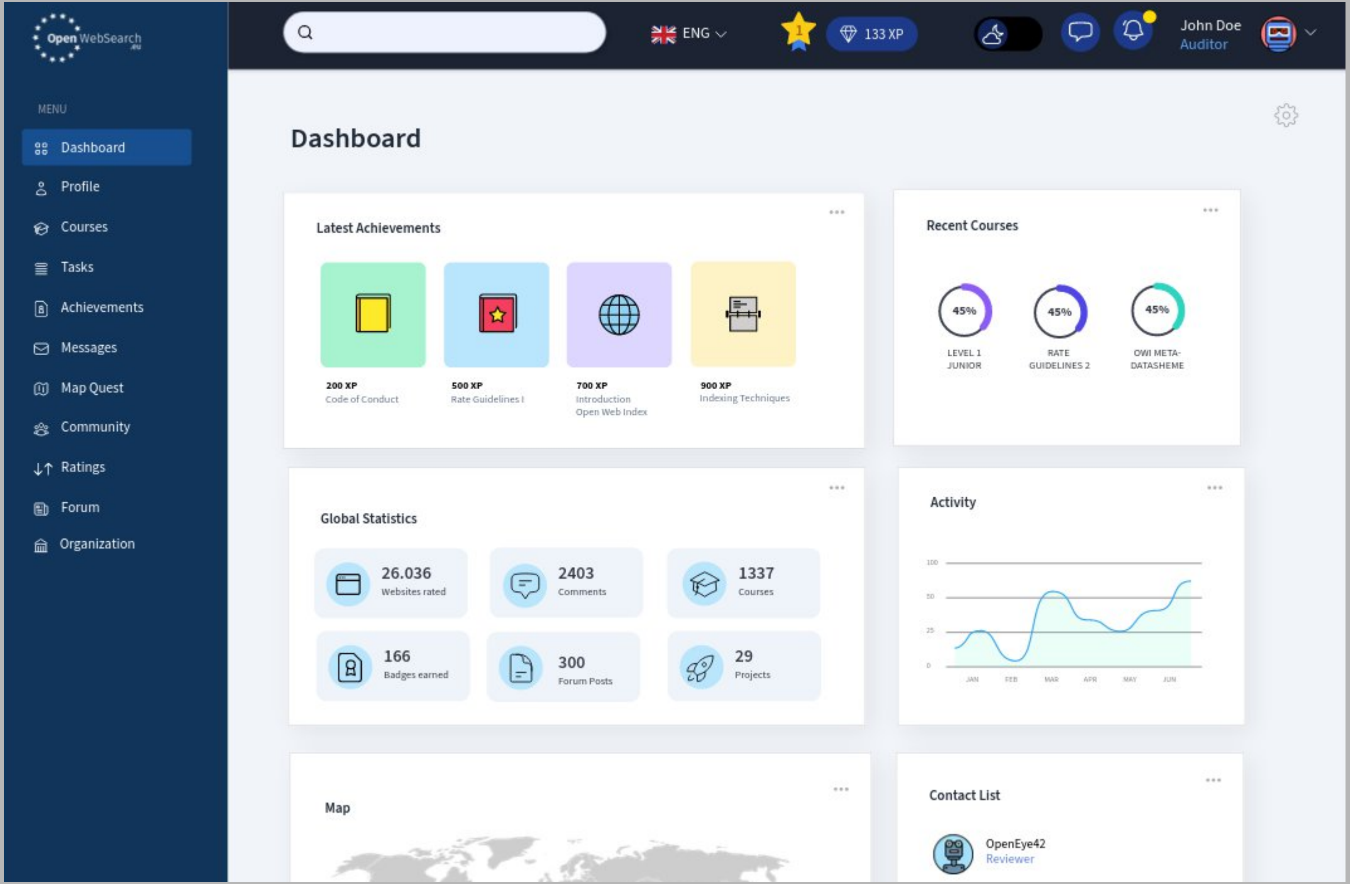
### Organization

Hierarchy of the level system:
Newbies
Mentors
Experts

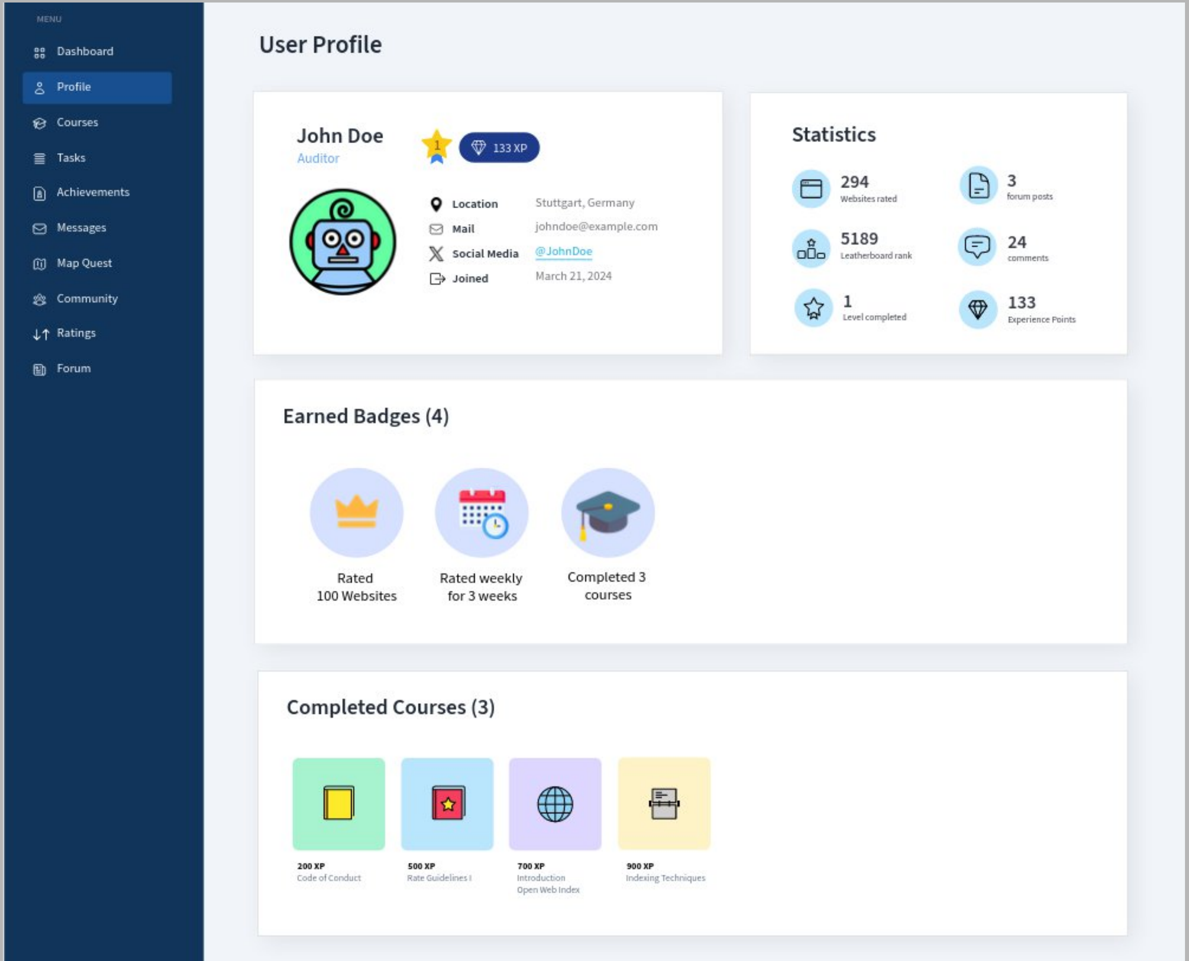
### Community Forum

Discussion and documentation of decision-making processes.

# Dashboard


Open WebSearch.eu
ENG
1
133 XP
John Doe
Auditor
MENU
Dashboard
Profile
Courses
Tasks
Achievements
Messages
Map Quest
Community
Ratings
Forum
Organization
Dashboard
Latest Achievements
200 XP
Code of Conduct
500 XP
Rate Guidelines I
700 XP
Introduction Open Web Index
900 XP
Indexing Techniques
Recent Courses
45%
LEVEL 1 JUNIOR
45%
RATE GUIDELINES 2
45%
OWI META-DATASHEME
Global Statistics
26.036
Websites rated
2403
Comments
1337
Courses
166
Badges earned
300
Forum Posts
29
Projects
Activity
100
50
25
0
JAN
FEB
MAR
APR
MAY
JUN
Map
Contact List
OpenEye42
Reviewer

# User Profile



## User Profile

John Doe
Auditor

| | |
|---|---|
| Location | Stuttgart, Germany |
| Mail | johndoe@example.com |
| Social Media | @JohnDoe |
| Joined | March 21, 2024 |

### Statistics

| | |
|---|---|
| 294 Websites rated | 3 forum posts |
| 5189 Leatherboard rank | 24 comments |
| 1 Level completed | 133 Experience Points |

### Earned Badges (4)

Rated 100 Websites

Rated weekly for 3 weeks

Completed 3 courses

### Completed Courses (3)

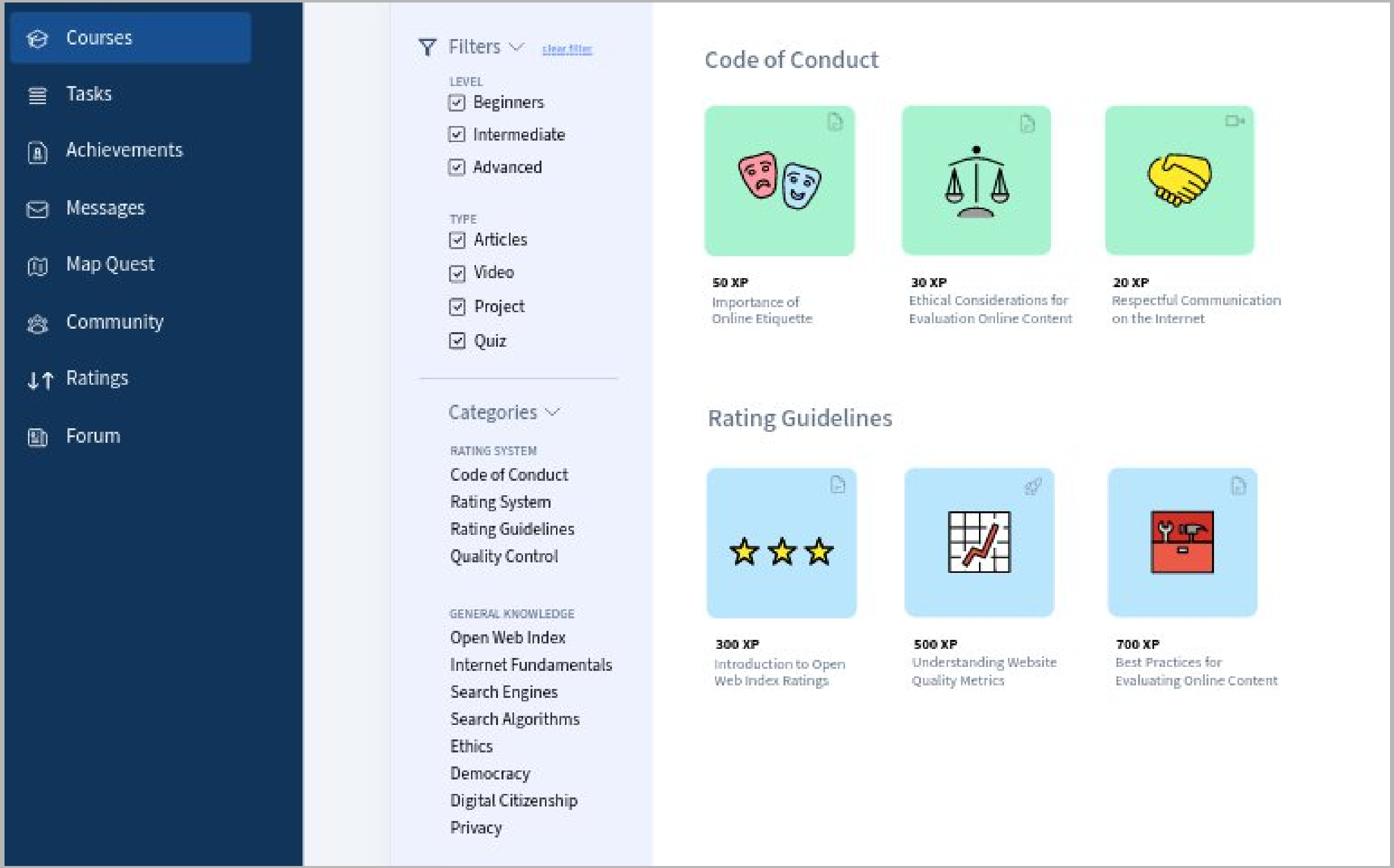
**200 XP** Code of Conduct

**500 XP** Rate Guidelines I

**700 XP** Introduction Open Web Index

**900 XP** Indexing Techniques

# Courses



LEVEL

- ☑ Beginners
- ☑ Intermediate
- ☑ Advanced

TYPE

- ☑ Articles
- ☑ Video
- ☑ Project
- ☑ Quiz

RATING SYSTEM

- Code of Conduct
- Rating System
- Rating Guidelines
- Quality Control

GENERAL KNOWLEDGE

- Open Web Index
- Internet Fundamentals
- Search Engines
- Search Algorithms
- Ethics
- Democracy
- Digital Citizenship
- Privacy

## Code of Conduct

**50 XP**
Importance of Online Etiquette

**30 XP**
Ethical Considerations for Evaluation Online Content

**20 XP**
Respectful Communication on the Internet

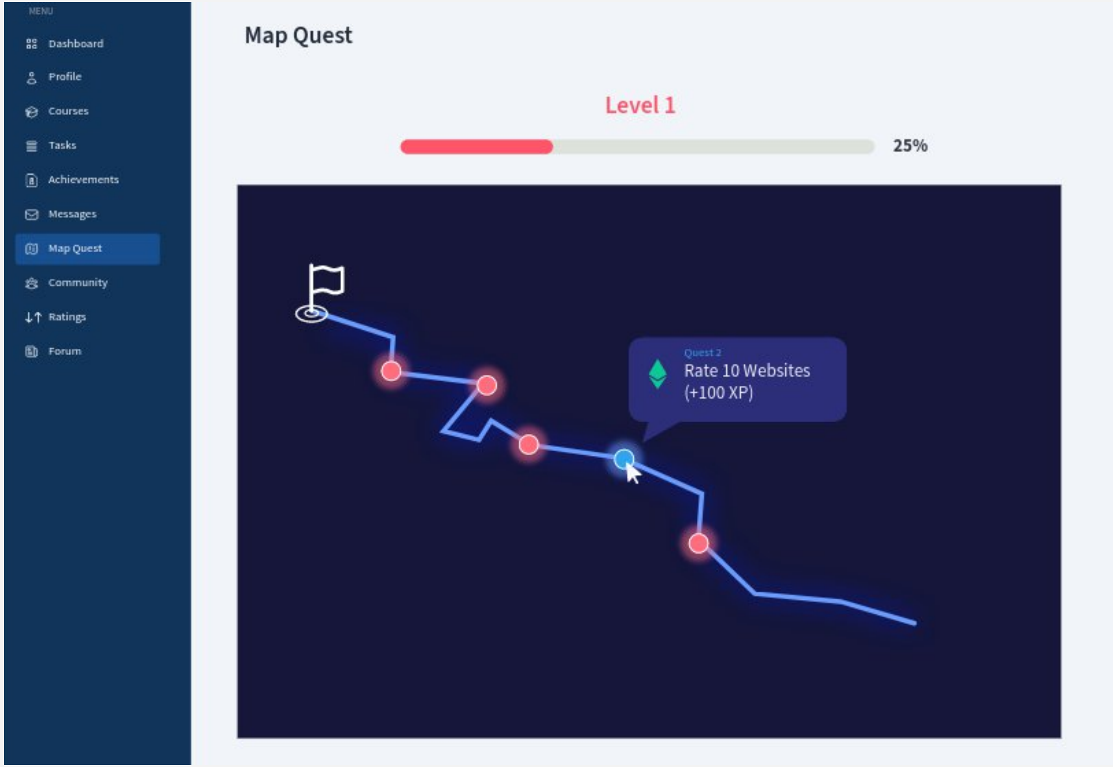
## Rating Guidelines

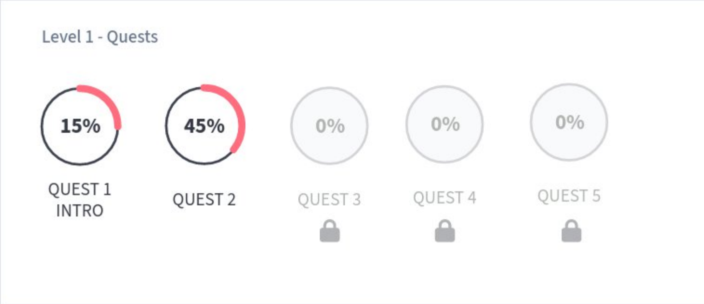
**300 XP**
Introduction to Open Web Index Ratings

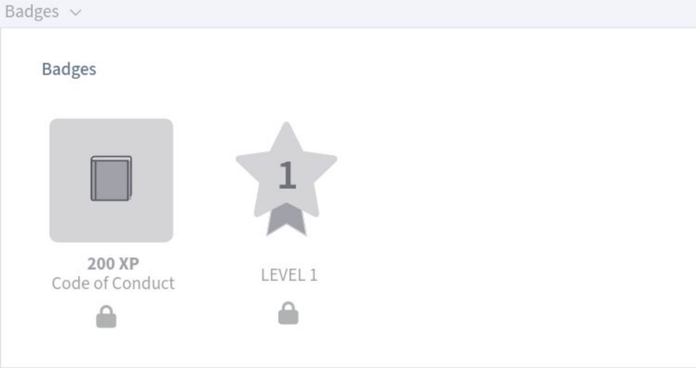
**500 XP**
Understanding Website Quality Metrics

**700 XP**
Best Practices for Evaluating Online Content

# Quest Map


MENU
Dashboard
Profile
Courses
Tasks
Achievements
Messages
Map Quest
Community
Ratings
Forum
Map Quest
Level 1
25%
Quest 2
Rate 10 Websites
(+100 XP)

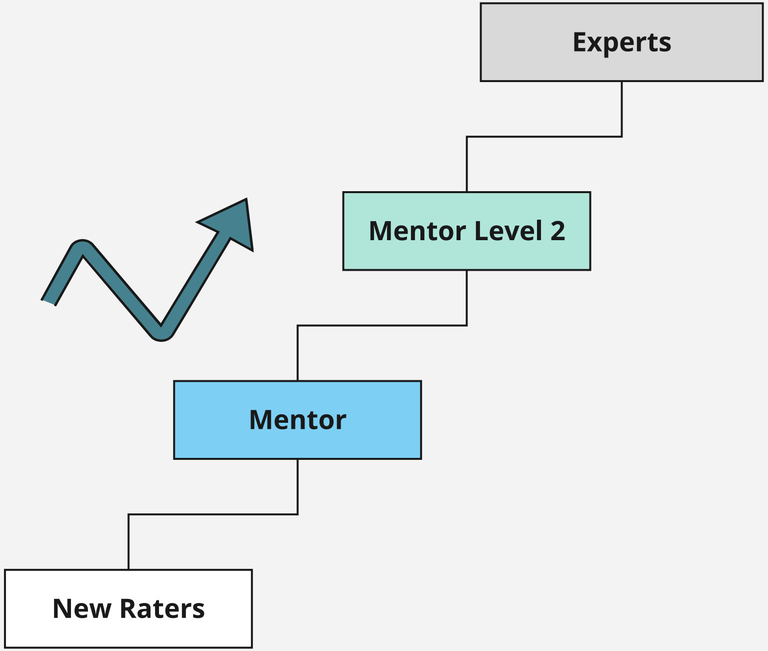

Level 1 - Quests
15%
QUEST 1
INTRO
45%
QUEST 2
0%
QUEST 3
0%
QUEST 4
0%
QUEST 5


Badges
Badges
200 XP
Code of Conduct
1
LEVEL 1

# Level up

By completing courses, users could gain more in-depth information on specific subject areas.

Users can earn badges that are displayed on their profile, which then entitles them to rate websites in specific subject areas, similar to the Dewey Decimal Classification system.



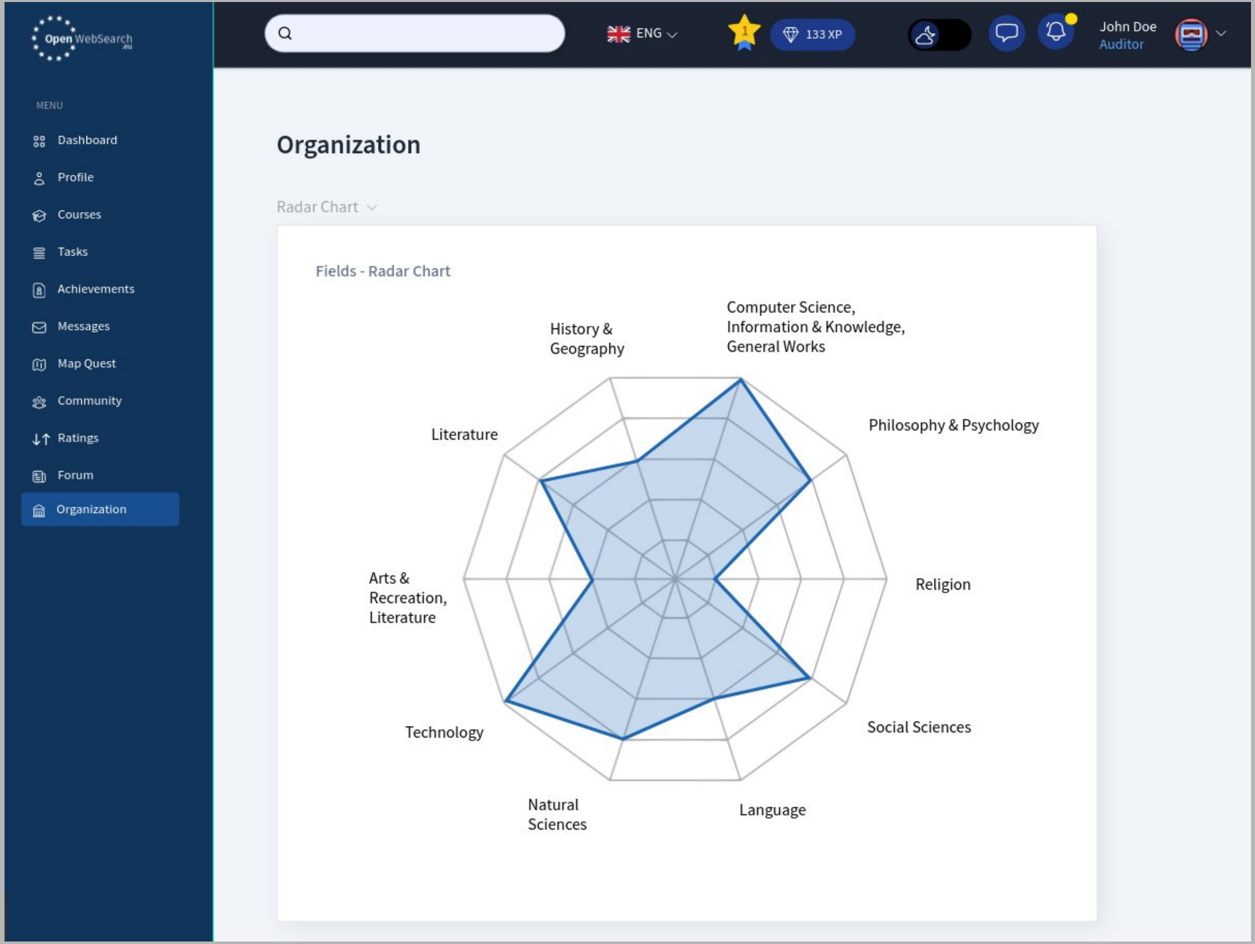
Experts

Mentor Level 2

Mentor

New Raters

# Experts


Open WebSearch.eu
ENG
1
133 XP
John Doe
Auditor
MENU
Dashboard
Profile
Courses
Tasks
Achievements
Messages
Map Quest
Community
Ratings
Forum
Organization
Organization
Radar Chart
Fields - Radar Chart
Computer Science, Information & Knowledge, General Works
Philosophy & Psychology
Religion
Social Sciences
Language
Natural Sciences
Technology
Arts & Recreation, Literature
Literature
History & Geography

# Rating Interface

Info →

Tools →

Rating →

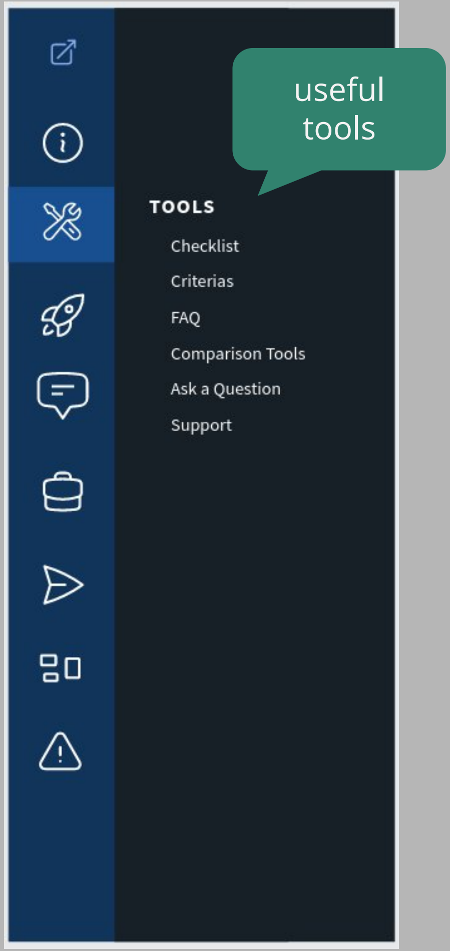
Comment →

Expert →

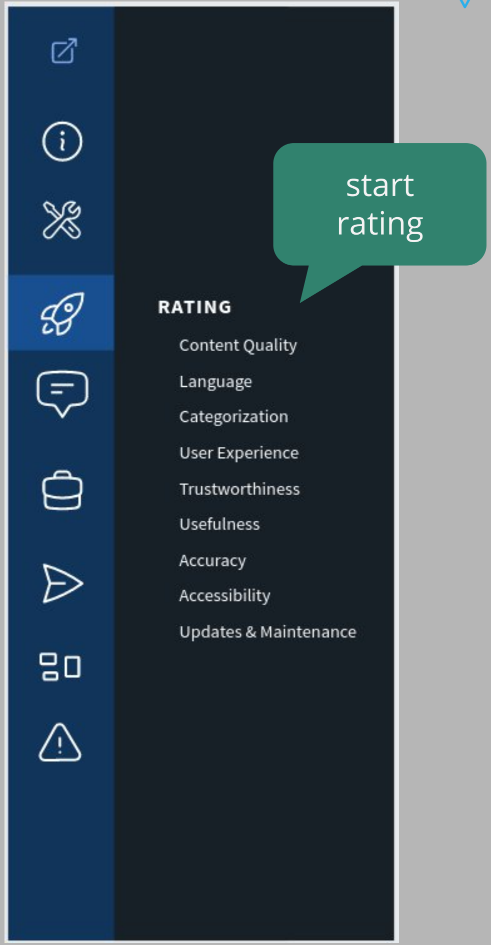
Feedback →

Sections →

Report →



useful tools

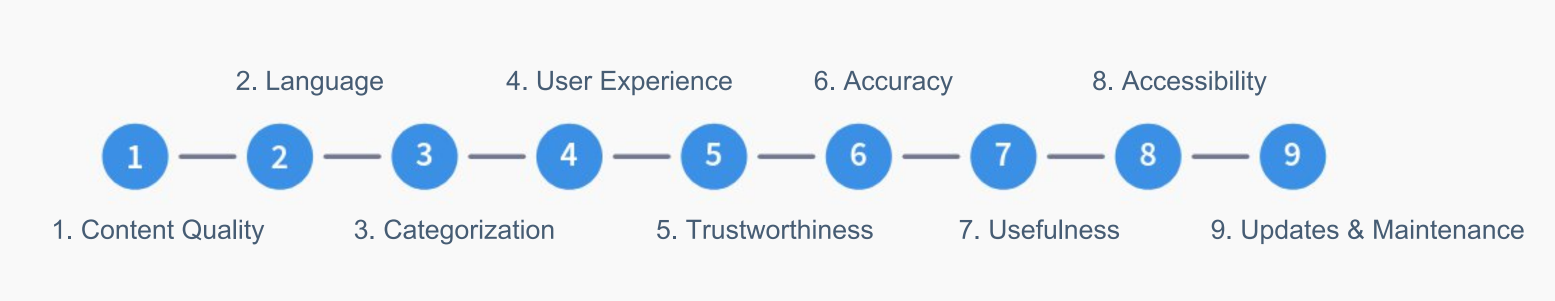


start rating

# Rating Interface

easy walkthrough

## step by step



1. Content Quality

2. Language

3. Categorization

4. User Experience

5. Trustworthiness

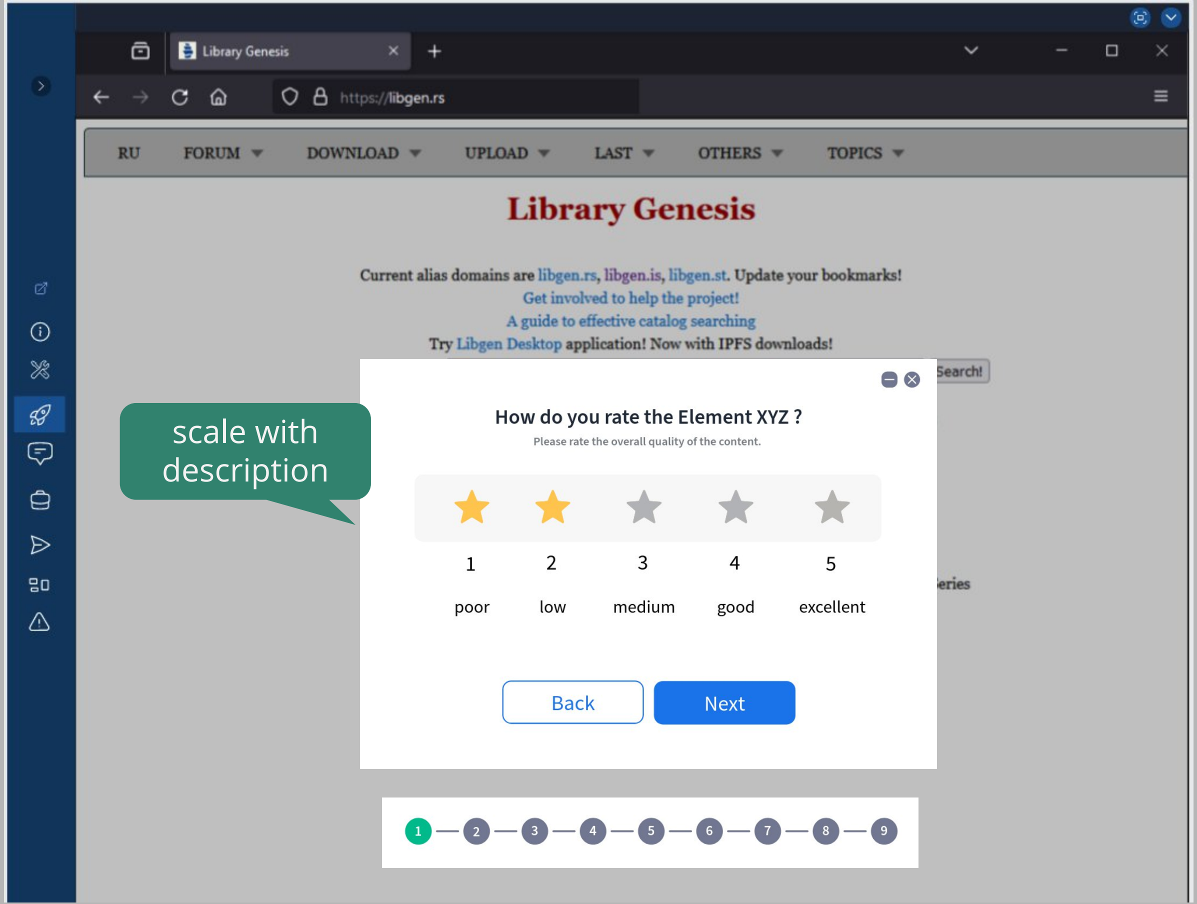
6. Accuracy

7. Usefulness

8. Accessibility

9. Updates & Maintenance

# Rating Interface



Library Genesis

Current alias domains are libgen.rs, libgen.is, libgen.st. Update your bookmarks!
Get involved to help the project!
A guide to effective catalog searching
Try Libgen Desktop application! Now with IPFS downloads!

scale with description

How do you rate the Element XYZ ?

Please rate the overall quality of the content.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| poor | low | medium | good | excellent |

# Preventing Negative Aspects

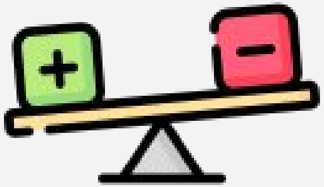
# Key Principles for Our Rating Platform

- Code of Conduct
- Democratic Participation & Self-Determination
- Welcoming Culture, Diversity & Inclusion
- Discussion Culture & Critical Engagement
- Fair Treatment & No Discrimination
- Open Source Philosophy & Privacy-Friendly Tools
- User Verification and Authentification for increased responsibility

# Implementing Majority Decision in Reviews

- Peer-Review System
- -> majority decisions and expert review
- Reviews are not shown individually but as part of an aggregated score or decision
- outlier reviews (extremely positive or negative) are identified and discussed

# Advantages of Majority Decision in Reviews



- Balanced Perspective: Reduces the impact of extremely biased reviews.
- Increased Credibility: Aggregated reviews are often seen as more trustworthy by users.
- Mitigation of Extremes: Helps in mitigating the impact of outlier opinions and potential fake/bad reviews.

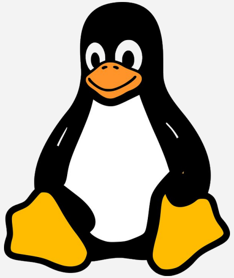
# Motivation

# Role Models for Community Engagement

## Wikipedia

https://pixabay.com/illustrations/wikipedia-books-encyclopedia-1802614/

## Open Source Community



https://pixabay.com/vectors/linux-unix-tux-penguin-cute-2025130/

## Chaos Computer Club

Photo by Susanne Krol

## Cryptoparties

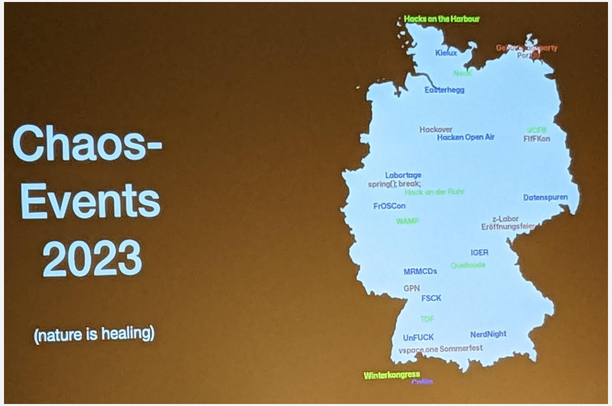
https://github.com/cryptoparty/flyers/blob/master/CryptoParty-Berlin/2016/berlin_2016_de_long_105_210.pdf

## Open Street Map

https://en.wikipedia.org/wiki/OpenStreetMap#/media/File:Openstreetmap_logo.svg liscence: CC BY-SA 3.0

# Example: Chaos Computer Club

- registered association
- decentrally organized in regional groups
- regional branches
- local associations with club rooms
- famous speakers
- media library with recordings of lectures
- workshops
- annual event Chaos Communication Congress, Chaos Camp
    - many other smaller events distributed throughout Germany



Die Einhörner freuen sich über Futter :)
CYBER
Making paper boats
make your own button with your own design

Photos by Susanne Krol

# Example: Chaos Computer Club

## Culture:

- "Hacker Bible"
- Magazine “Datenschleuder”
- Chaosradio (Podcast)

## Society:

- Commitment to freedom of information, data protection
- Appearing as experts, submitting statements
- "Chaos makes school" - Cooperation with educational institutions

Die
TEIL 1
HACKERBIBEL
Chaos Computer Club
URGH!
ROBIN HOOD
IM DATEN-NETZ
KABELSALAT
IST GESUND!
GRÜNER ZWEIG 98

Photos by Susanne Krol

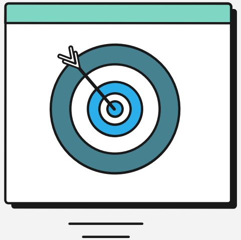
# Create & Promote Culture

## Personal meetups:

- Local Circles
- Association Work
- Meet-ups
- Annual Event

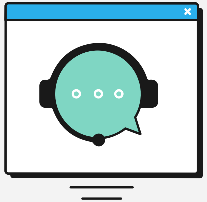
## Strategy:

- Intrinsic Motivation
- Gamification
- Mascot (Merch)
- Media Attention
- Creating Awareness
- Mentorship Program
- Knowledge Transfer
- Media Library, E-Learning

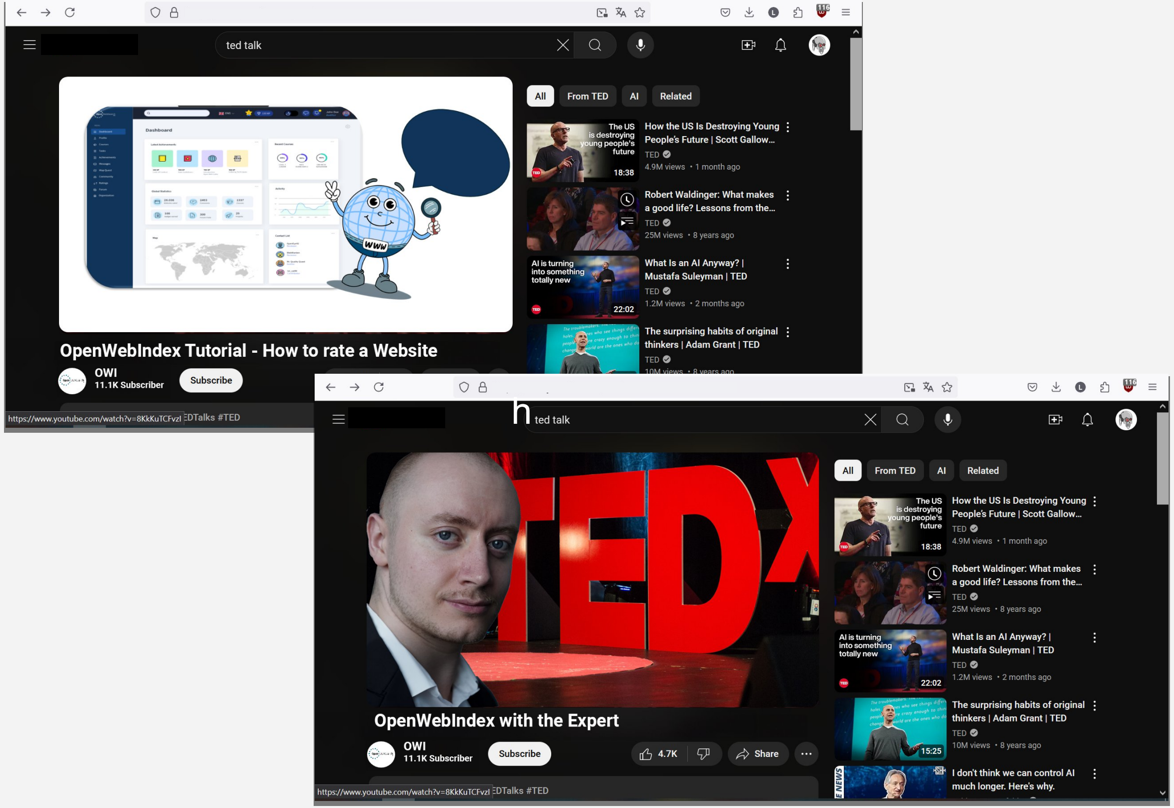


## Digital Exchange:

- Mailing Lists
- Social Media Integration / Fediverse
- Community Forum
- Matrix Channels (Chat/Messenger)
- BigBlueButton Rooms



# Media Library for the OWI



[redacted]

ted talk

All | From TED | AI | Related

How the US Is Destroying Young People's Future | Scott Gallow…
TED
4.9M views • 1 month ago

Robert Waldinger: What makes a good life? Lessons from the…
TED
25M views • 8 years ago

What Is an AI Anyway? | Mustafa Suleyman | TED
TED
1.2M views • 2 months ago

The surprising habits of original thinkers | Adam Grant | TED
TED
10M views • 8 years ago

OpenWebIndex Tutorial - How to rate a Website
OWI
11.1K Subscriber
Subscribe

https://www.youtube.com/watch?v=8KkKuTCFvzI EDTalks #TED

[redacted]

h ted talk

All | From TED | AI | Related

How the US Is Destroying Young People's Future | Scott Gallow…
TED
4.9M views • 1 month ago

Robert Waldinger: What makes a good life? Lessons from the…
TED
25M views • 8 years ago

What Is an AI Anyway? | Mustafa Suleyman | TED
TED
1.2M views • 2 months ago

The surprising habits of original thinkers | Adam Grant | TED
TED
10M views • 8 years ago

I don't think we can control AI much longer. Here's why.

OpenWebIndex with the Expert
OWI
11.1K Subscriber
Subscribe
4.7K | Share

https://www.youtube.com/watch?v=8KkKuTCFvzI EDTalks #TED

The US is destroying young people's future
18:38
AI is turning into something totally new
22:02
15:25
Dashboard
www
TEDx

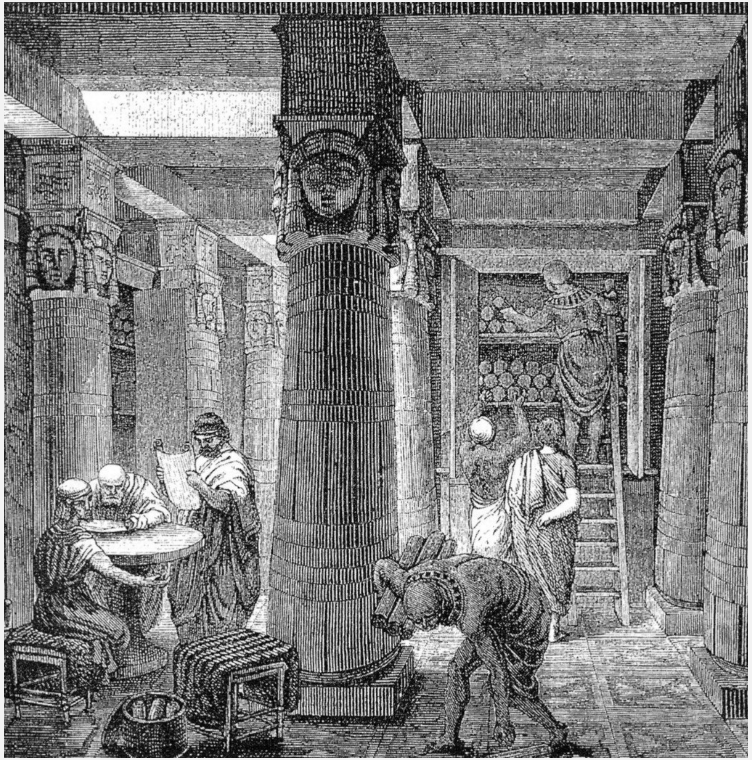
## Ideas:

- Animated explanatory videos
- TED Talks
- Behind the scenes
- Tutorials
- PeerTube as alternative to Youtube

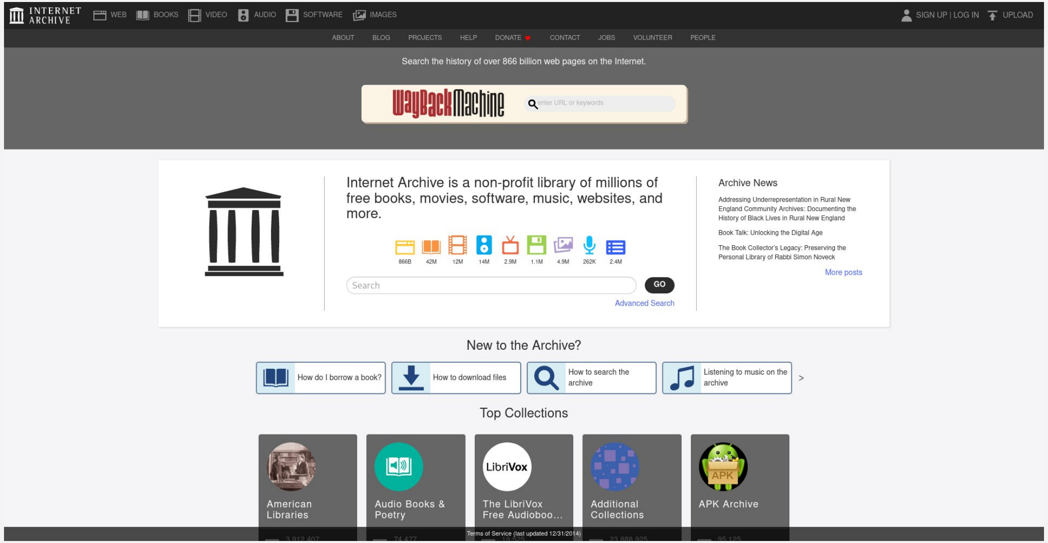
# Libraries
# &
# Open Web Index

# Librarians

## ...have always been experts in information management



The Library of Alexandria, license: public domain



Search the history of over 866 billion web pages on the Internet.

Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.

Archive News

Addressing Underrepresentation in Rural New England Community Archives: Documenting the History of Black Lives in Rural New England

Book Talk: Unlocking the Digital Age

The Book Collector's Legacy: Preserving the Personal Library of Rabbi Simon Noveck

New to the Archive?

Top Collections

Terms of Service (last updated 12/31/2014)

screenshot from https://archive.org/

# Libraries X Open Web Index

## What could a collaboration look like?

https://www.pexels.com/photo/librarian-at-work-11885954/

**Events:**

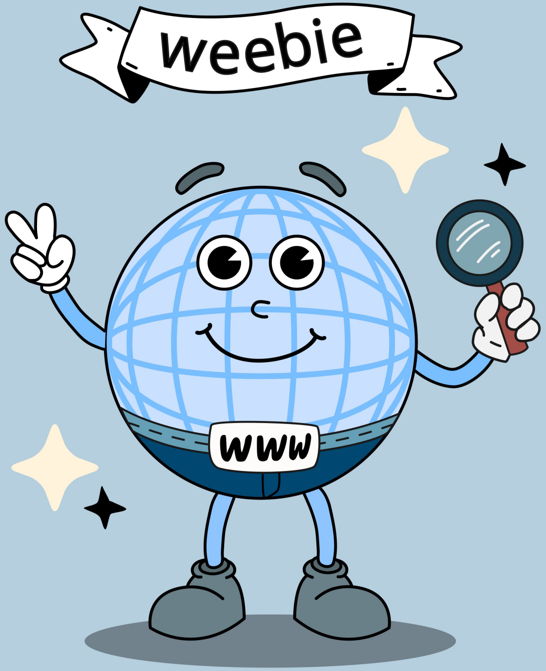
- hosts for events
- local clubs with meetups
- collaboration with universities

https://www.pexels.com/photo/people-sitting-on-chair-in-front-of-table-4865515/

**Expertise from librarians:**

- Information organization and indexing
- Quality control
- Inventory development and management
- Promotion of information literacy
- Public relations and visibility

# OWI Mascot



# Thanks for listening!