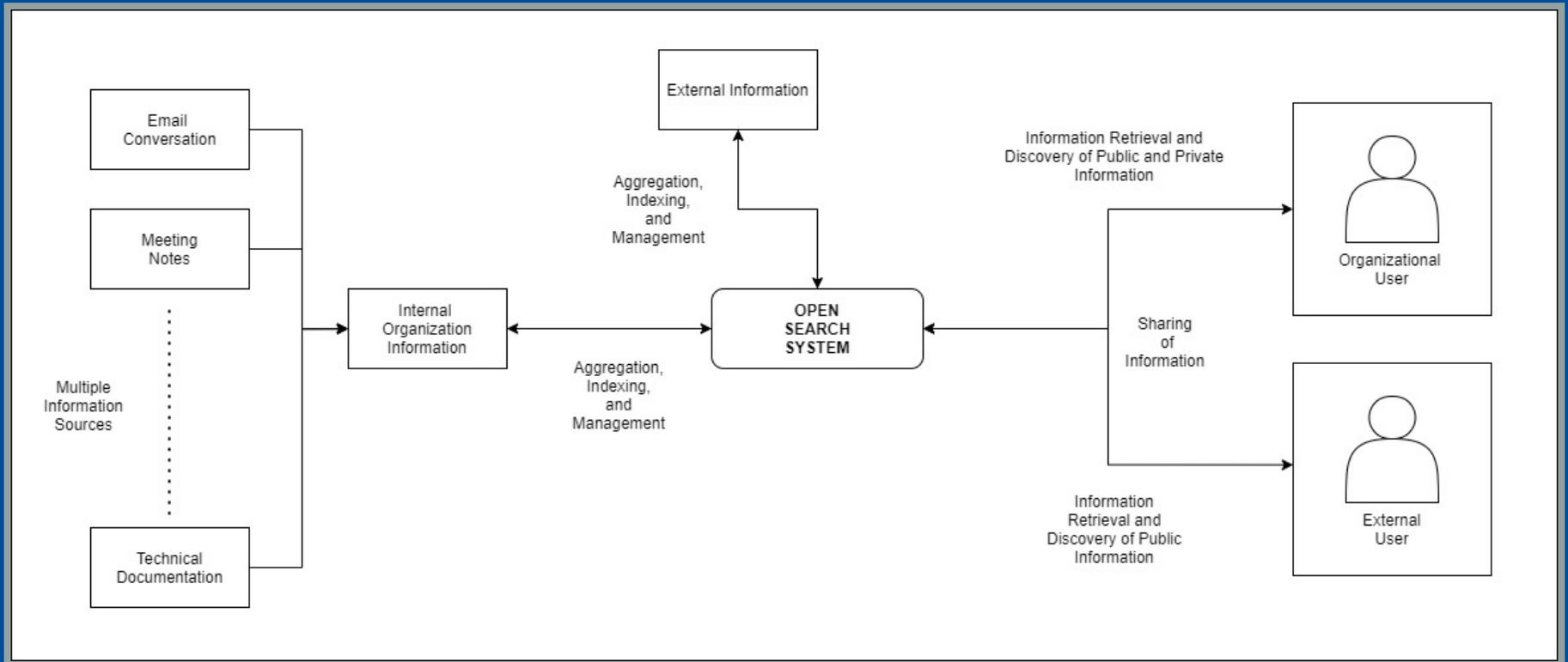


A Proposal for Client Based User Profiles For Open Search in Large and Highly Connected Organizations

DI Igor Jakovljevic, CERN & Graz University of Technology
Prof. Dr. Christian Guetl, Graz University of Technology
Dr. Andreas Wagner, CERN

Introduction

- Humans generate 33 trillion gigabytes of information per minute
- Organizations process 500 TB of data per week
- As the amount of data produced by humans on the Web and within large organizations also rapidly increases, new challenges for navigation through information are formed
- At the 2nd OSSYM an open information-based search system was presented



Introduction

- The purpose of this system was to offer large organizations the ability to share information transparently, enabling information retrieval to organizational users and also external users, while offering a high degree of data protection and privacy to users.
- The system offered a high level of privacy and data protection by not tracking user behavior and not invading user-sensitive information, which resulted in the system lacking a level of user personalization.

Personalization

- For search engines to adjust the results of queries, the search engines must :
 - Know which user sent the query
 - the personal information of the user that might benefit the result
 - understand the context of the query.
- This information about the user can be collected explicitly by asking the user to provide the information or implicitly by collecting user behavior data.

Personalization

- Retaining user search query logs, search engine service providers can provide additional services:
 - enhancing ranking algorithms
 - query fine-tuning
 - improving personalized query results
 - combating fraud and abuse
 - using data for research
 - using data for marketing and other commercial purposes

Personalization

Personalization information is often stored on the servers of the search engine providers.

Personalization Downsides

The downsides of preserving metadata and user data can lead to serious privacy problems as well.

This allows these providers to exploit user information for monetary gain, opportunities to steal personal information, and general misuse of information

The keywords of each query and the related metadata may disclose sensitive user information such as behaviors, habits, interests, religious views, sexual orientation, etc.

Personalization Downsides Example

In the year 2006, the internet company AOL released a large amount of user search requests to the public for research purposes. The information was anonymized and did not contain any user information, but personally identifiable information was present in many of the queries. This enabled users to be identified by their search histories

Personalization

Less than 40% of Europeans are comfortable with the idea of search engine providers accessing their online activity to improve advertising or content

Eurobarometer survey on Data Protection and Electronic Identity in the European Union, 2011

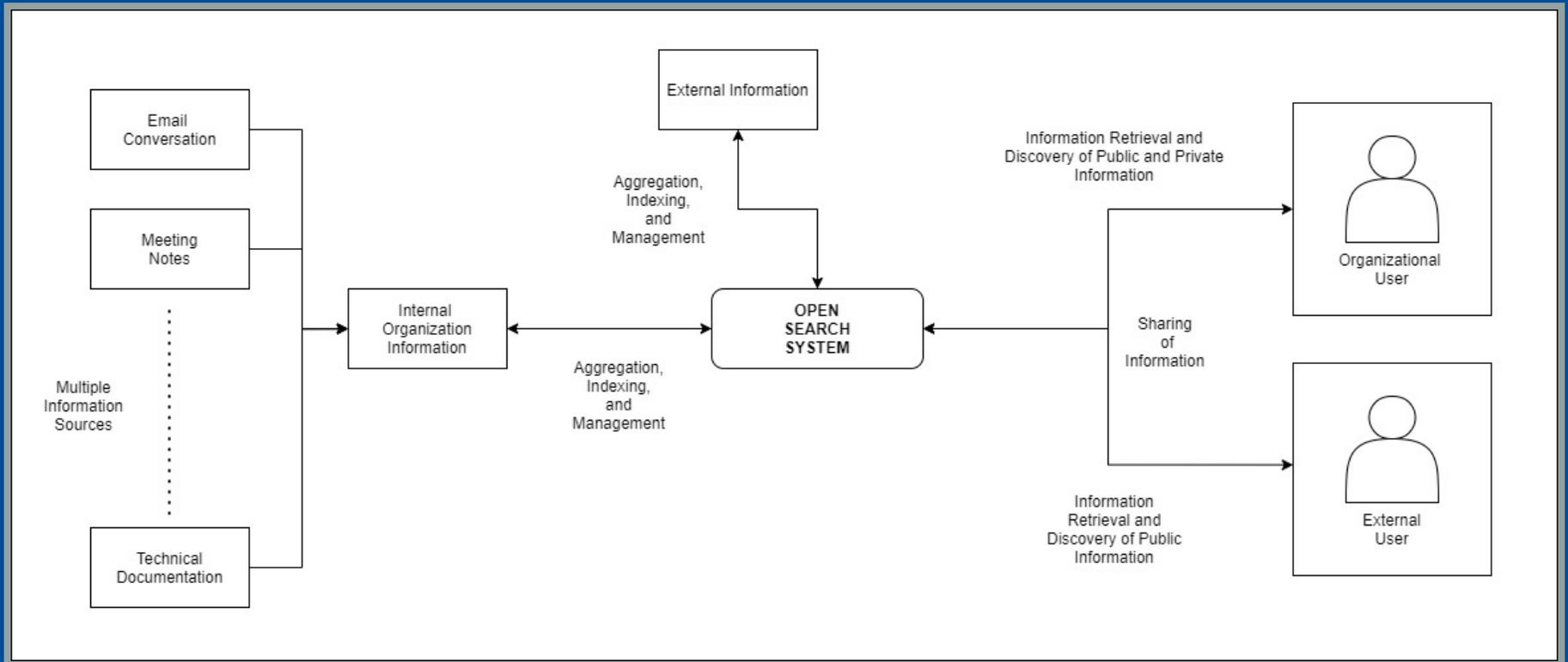
Focus of the Paper

Based on the analysis of different search engine systems, user profiling methods, and literature survey on information sharing in large organizations

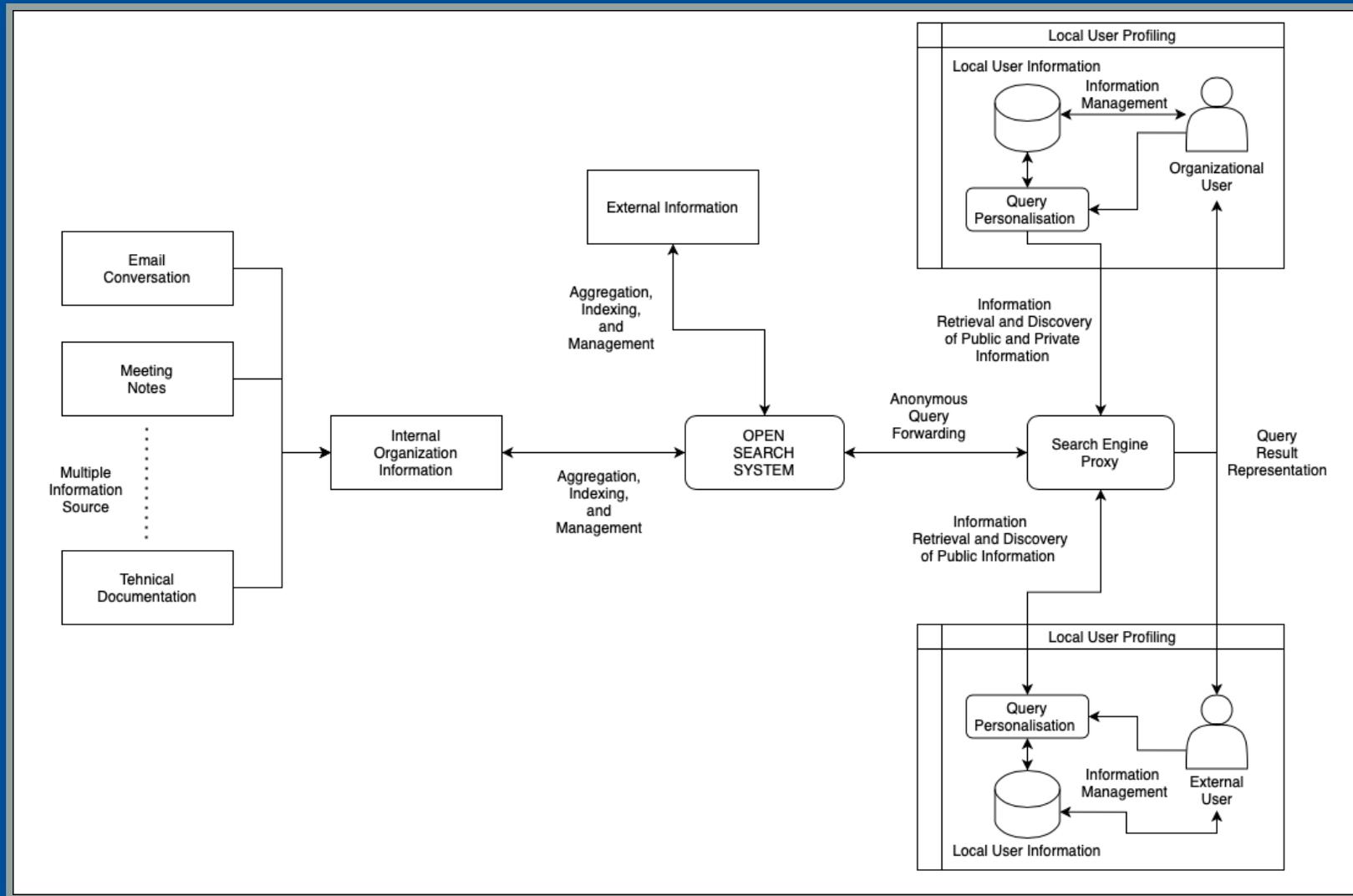
we want to introduce

a conceptual extension to the open information-based search system that integrates aspects of personalization, data privacy and protection into an open search architecture

TOWARDS A CLIENT BASED SEARCH



TOWARDS A CLIENT BASED SEARCH



Updated Conceptual Integration Diagram of the Open Search System for Large and Highly Connected Organizations

TOWARDS A CLIENT BASED SEARCH

- Creation and Maintenance of User Profiles
- Local Query Personalisation
- Data Protection and Search

TOWARDS A CLIENT BASED SEARCH

Creation and Maintenance of User Profiles

- Store user profile information within the user client, within the browser database or with the use of a plugin to store it on the file system
- Keep track of all these information sources and aggregate them into one coherent user profile, while securely sharing it among multiple user devices and services

TOWARDS A CLIENT BASED SEARCH

Local Query Personalisation

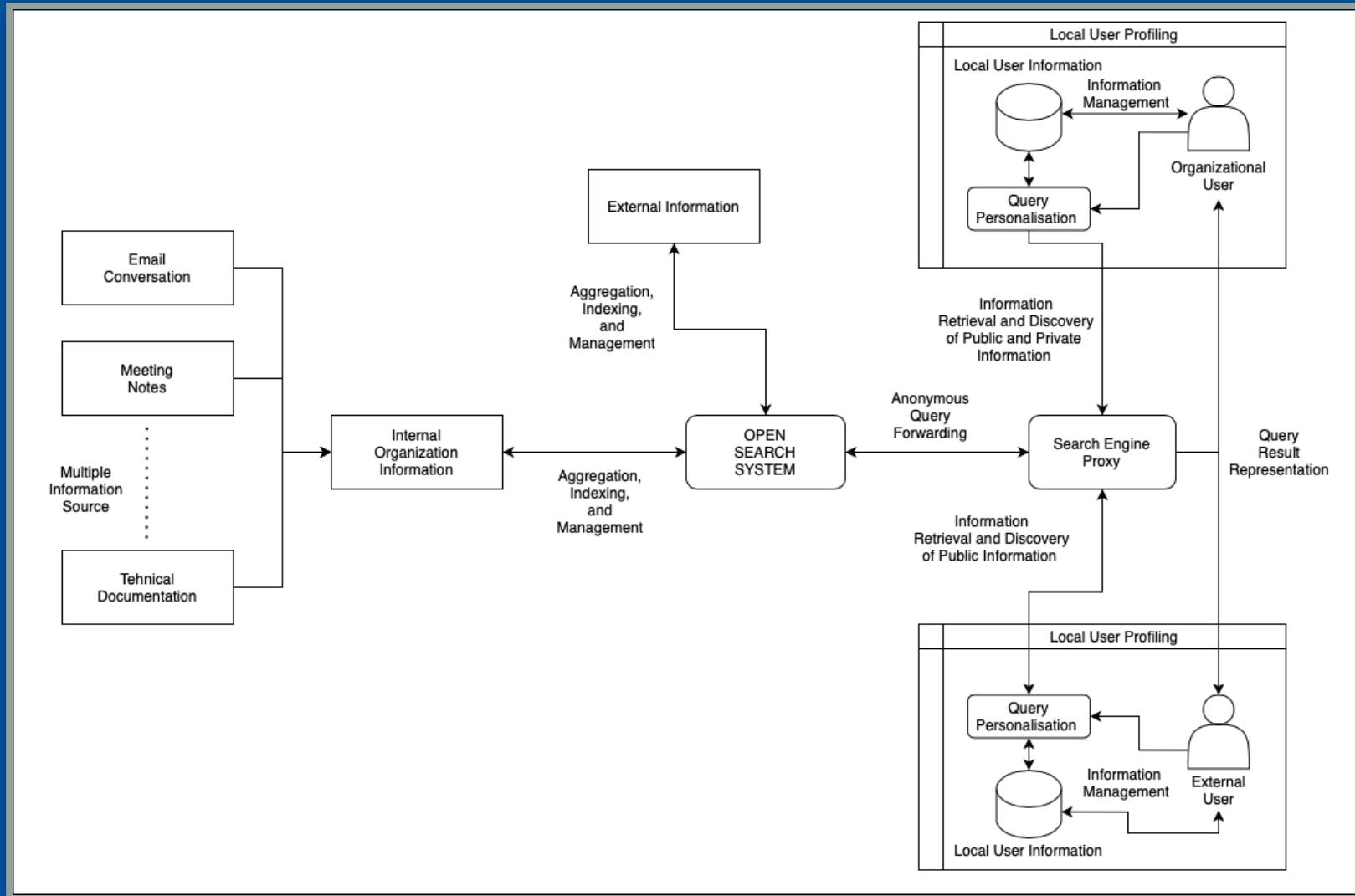
- Query personalization component is used to enhance a user query into a personalized user query
- Creating Personalized Queries Locally

TOWARDS A CLIENT BASED SEARCH

Data Protection and Search

- User profiles are securely shared between multiple devices using a local blockchain, where each user would create a local blockchain network, which would be updated as soon as one device updates a user profile
- Another implementation could include the usage of Solid Pods
- To increase the level of privacy of user data, we propose the integration of a search engine proxy

TOWARDS A CLIENT BASED SEARCH



Updated Conceptual Integration Diagram of the Open Search System for Large and Highly Connected Organizations

Conclusion

- Technical challenges vary from efficiently storing, generating, maintaining, and sharing user profiles, using those profiles for secure and effective personalization and information retrieval
- Issues with user privacy and user rights need to be analyzed in-depth to find new ways to educate the user about the data that is being collected and processed

Thank you for your Attention

Contacts

- Twitter: @IgoJJ
- Email:
 - igor.jakovljevic@outlook.com
 - igor.jakovljevic@cern.ch

