



**Fraunhofer**

IAIS

FRAUNHOFER INSTITUTE FOR INTELLIGENT  
ANALYSIS AND INFORMATION SYSTEMS IAIS

**FuhSen**

A SEMANTIC FEDERATED HYBRID  
SEARCH ENGINE



**FuhSen**

How can we generate value from big data on the Internet in a legally safe way? Overcome the challenging heterogeneity of information about persons, products and organizations with FuhSen – The Semantic Federated Hybrid Search Engine!

## GLOBAL CHANGE

Today, more data races through the Internet every second than was saved in total over the past 20 years. We are walking data producers – social media activities, the use of cloud services and self-tracking via the smartphone have brought about a growth in databases every day, hour and second. Be it surfing online or out and about with a smartphone, humans leave a permanent trail of data that makes it possible to draw a detailed picture of them. Alongside demographic data, orders of products and services, payment transactions, location and social contacts are increasingly flowing into this profile. This allows devices and services to get to know their users in detail and to personalize the user experience.

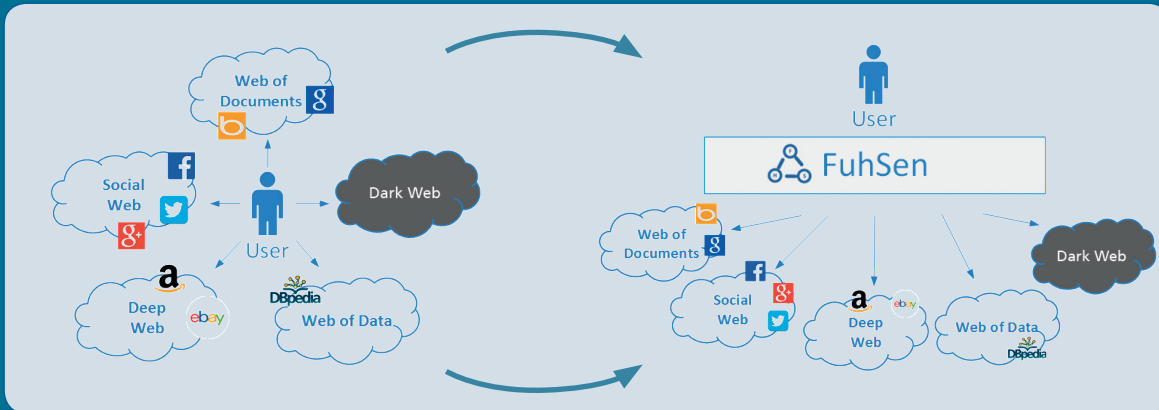
With the help of big data techniques, companies have started to comb their way through the wealth of their customer data in search of hidden patterns of valuable information. How can we generate value from data that has not been collected over years, but that exists, or is being generated on the Web right now, in a legally safe way? This question is addressed by Fraunhofer IAIS.

## IDENTIFIED CHALLENGES

Headhunters and HR officers spend much time gathering information searching a variety of websites and networks for the skills and qualifications of a candidate for a job. As another example, criminal investigators have a hard time collecting all information about a suspected person or organization, or about an illegal product offer, and identifying the connections between these (e.g., who is trading what).

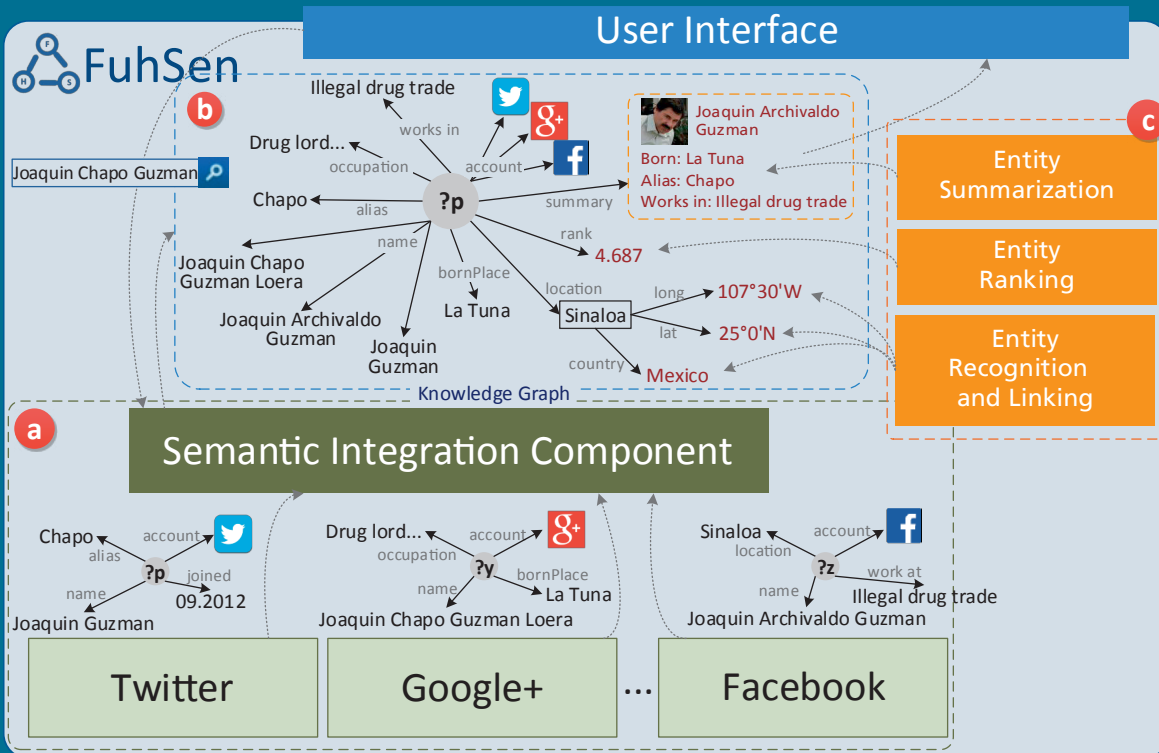
A lot of the required information is available on the Web, e.g., in social networks, but scattered over multiple different sites. The sources of this information change frequently, and because of legal restrictions concerning data protection and privacy, it is not allowed and desirable to store results of such research unconditionally. This requires on-demand search and information integration. Still, for each search result its provenance has to be transparent, i.e., at what source it has been found. Different sources provide information in different data formats and schemas and have different accessibility restrictions; this heterogeneity of data needs to be considered.

Despite the complexity of the data, the search user interface has to be learnable for its users, who are experts in HR or criminal investigation but should not have to be expert data engineers.



Chaotic search on different platforms

Structured & simplified search on one platform



Architecture of FuhSen: a) Components that extract and integrate the data on-demand from the data sources b) Knowledge Graph of results c) Data enrichment using state-of-the-art semantic algorithms

## LEAD WITH FuhSen

Following the input of a keyword, the FuhSen search engine collects on demand data spread over different sources on the Web (Social Networks, e-Commerce Platforms, Open Data). The data is integrated into a coherent knowledge graph, results can be filtered, visualized and displayed as short summaries.

Via its semantic integration component FuhSen aggregates information about people, organizations and products, and integrates pieces of information on demand. Furthermore, FuhSen enriches entity data applying state-of-the-art semantic technology such as entity recognition, linking, ranking and summarization.

FuhSen provides a keyword search interface. Results are grouped by entity type, and filters are generated automatically based on the data to explore the results. The user can visualize the results in a list, graph or map. Results can also be exported for external analysis.

### Application Areas

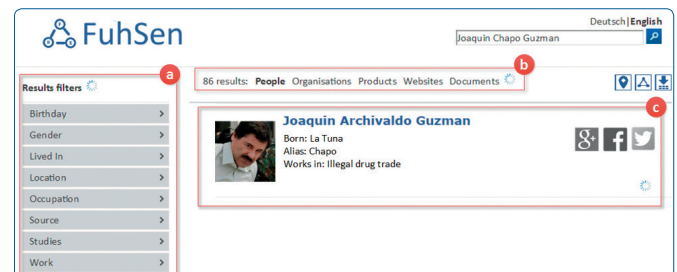
- Criminal analysis for law enforcement agencies
- Job posting search
- Identification and CV overview of skilled professionals

### Reference Project

- LiDaKrA: Integration of linked information and early detection of organized crime, research project funded by the Federal Ministry of Education and Research

### Publication

- Diego Collarana, Christoph Lange, Sören Auer: »FuhSen: A Platform for Federated, RDF-based Hybrid Search«. In: WWW'16 Companion. 25<sup>th</sup> International World Wide Web Conference. ACM, 2016, pp. 171–174.



User Interface in FuhSen: a) Filters to explore the results b) Results are grouped by type c) A summary of the entity data is displayed



## **Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS**

Schloss Birlinghoven  
53757 Sankt Augustin  
Germany

[www.iais.fraunhofer.de](http://www.iais.fraunhofer.de)

### **Enterprise Information Systems**



**Dr. Christoph Lange-Bever**

Expert Group Manager »Data Integration«

Phone: +49 2241 14-2428

[christoph.lange-bever@iais.fraunhofer.de](mailto:christoph.lange-bever@iais.fraunhofer.de)



**Richa Sharma**

Business Developer & Project Manager

Phone: +49 2241 14-2303

[richa.sharma@iais.fraunhofer.de](mailto:richa.sharma@iais.fraunhofer.de)

[FuhSen@iais.fraunhofer.de](mailto:FuhSen@iais.fraunhofer.de)