FOM Hochschule für Oekonomie & Management

University Location Nürnberg

Bachelor Thesis

in the study course Wirtschaftsinformatik - Business Information Systems

to obtain the degree of

Bachelor of Science (B.Sc.)

on the subject

Design, Piloting, and Evaluation of an Expert Finder System Prototype for Enterprise Organizational Structures

by

Joschua Böhm

Advisor:	Prof. Dr. Klemens Waldhör
Matriculation Number:	604968
Submission:	2024-10-31

Contents

List of Figures

List of Abbreviations

EFS	Expert Finding System
ELS	Expertise Location System
NLP	Natural Language Processing
AI	Artificial Intelligence
UI	User Interface
ML	Machine Learning
LLM	Large Language Model
API	Application Programming Interface
FAQ	Frequently Asked Questions

1 Introduction

1.1 Background

In this section, the background of the thesis is presented. The section is divided into five parts. First the term Expert Finding System (EFS) is defined, followed by providing a historical overview of the developments in the field of EFSs. Afterwards the current state of EFSs is briefly discussed. Finally, the organizational structures and technological foundations are presented in the last two parts of the section.

1.1.1 Definition of Expert Finding Systems

Traditionally, an Expert Finding System (EFS) or Expertise Location System (ELS) is a tool or technology, that enables users to identify and locate subject-matter experts with the aim of acquiring or utilizing the expert's knowledge (M. T. Maybury 2006, p. 1). According to M. T. Maybury (ibid., pp. vii, 3), an EFS must fulfill certain key requirements:

- **Identification:** An EFS must be able to identify experts. This can be achieved through self-nomination by the expert, or the automated identification based on documents like publications (ibid., pp. vii, 3), (Stankovic et al. 2010, p. 3), expert communication or the analysis of expert activities.
- **Classification:** An EFS, based on a variety of different sources of evidence, must classify the specific type and level of expertise an expert possesses, and
- **Validation:** Assess the range and depth of an expert's professional knowledge and skills. This can be done by human validation in the form of assessing qualification evidence, or by automated user feedback in the form of ratings or reviews.
- **Ranking:** The system must have the ability to rank experts based on different factors like field of expertise, experience, certification, publications and reputation.
- **Recommendation:** An EFS has to be able to return an ordered list of experts or groups of experts based on specific expertise needs in combination with importance criteria (e.g., experience, reputation etc.).

While the traditional definition of EFS offers a comprehensive overview regarding the requirements for identification and classification of experts effectively, the specific context of this thesis requires some modifications for the definition to be applicable to the use case of the thesis. The thesis is less focused on the knowledge transfer between experts and users, but rather on a much more goal-oriented approach in which the main objective is to find specific contacts in the company in order to solve or complete certain problems or tasks. The focus is therefore less on learning what the experts know but more on the system telling the user who the experts are and what they need to complete certain tasks, or even providing the user with workflows to solve the task. Based on this, the following definition for EFS is proposed: An EFS is a specialized system, that not only identifies relevant experts and contact persons within an organization, but also provides the user with concrete recommendations for solving specific problems and providing the user with resources to solve the problems tasks efficiently. In contrast to the traditional definition, where transfer of knowledge is the main focus, the main goal of the EFS in this thesis is the direct support in accomplishing tasks.

1.1.2 Historical Context of Expert Finding Systems

First appearances of the concept of EFSs can be traced back to the 1990s and early 2000s with contributions like ‘Answer Garden’ by Ackerman and Malone (1990), *Enterprise expert and knowledge discovery* by Mattox, M. Maybury, and Morey (1999) and “Facilitating the Online Search of Experts at NASA Using Expert Seeker People-Finder” by Irma Becerra-Fernandez (2000) as well as ‘Facilitating the Online Search of Experts at NASA using Expert Seeker People-Finder’ by Irma Becerra-Fernandez (1999). These early systems had the primary goal of connecting users with experts in order to facilitate knowledge transfer and collaboration, often in an academic setting. (Mattox, M. Maybury, and Morey 1999, p. 1) (Irma Becerra-Fernandez 2000, pp. 3–3) (Irma Becerra-Fernandez 1999, p. 3). On the technical side, these systems relied heavily on keyword search with basic search approaches as can be seen in Mattox, M. Maybury, and Morey (1999, pp. 4–5) for example. However, early approaches of AI in the form of data mining and clustering techniques were utilized in Irma Becerra-Fernandez (2000, pp. 3–1) to enhance their functionality, and early research on the role of AI technologies in EFSs was conducted by Becerra-Fernandez (2000). Following, M. T. Maybury (2006) conducted extensive research on the topic of EFSs in the journal article “Expert Finding Systems”. Those early research efforts not only laid the foundation for the research on the topic, but also highlighted the potential, future technologies could have on the field of EFSs. Over the years, the need for EFSs has only increased, with expertise becoming more and more recognition as a key asset for companies (Husain et al. 2019, p. 1). Developments of the last 20 years like globalization, digitization or remote work and trends such as agile working have, on the other hand, added a new layer of complexity in this field, as people more often work from different locations and in different teams, and colleagues are spread across different departments and locations. This makes it increasingly difficult for employees to keep track of who the experts are in the company and how to contact them.

1.1.3 Current State of Expert Finding Systems

Today, EFSs are used in a variety of different contexts and industries like in academia, enterprise or medicine. Use cases range from finding research collaborators, over recommending developers for specific tasks, up to forming teams (Husain et al. 2019, pp. 2, 9). Additionally, the technologies used in EFSs have evolved over the years. While traditional EFSs relied heavily on keyword search as seen in Mattox, M. Maybury, and Morey (1999, pp. 4–5), modern EFSs utilize AI technologies like Natural Language Processing (NLP) and Machine Learning (ML) to enhance their search quality as well as information retrieval (Husain et al. 2019, pp. 19–20). Those technologies also enable the processing of significantly larger amounts of data resulting in a more effective counter to the data overload problem, which is one of the reasons EFSs are needed in the first place (ibid., p. 1).

1.1.4 Organizational Structures

Corporations face the challenge of implementing EFSs that fit their specific needs and organizational structures. Four of the most commonly used organizational structures are the functional structure, which focuses of a clear chain of command and separates the organization into different departments based of their expertise (*What is Matrix Organization?* 2024), a product- or market-based structure where different departments are based on different products or markets instead of expertise, the geographical structure which divides teams, based on their location and a process-based structure which groups the employees into teams based on the business processes they are engaging in. (Organ 2023) An alternative approach is the matrix structure. The matrix structure is on the rise with 84% of employees being “matrixed” in some way according to a study of cross-functional teams conducted by Gallup (Inc 2024, page 65). The Matrix organization stands out by having multiple lines of reporting, meaning that employees have two or more bosses effectively (*What is a Functional Structure in an Organization?* 2024; Organ 2023). This makes the Matrix organization a great match for agile working and cross-functional teams. The main challenge that needs to be addressed regarding an EFS in a Matrix organization are dynamic and constantly changing tasks and fields of expertise, as people are incentivized to grow in those environments. Therefore, an EFS has to be able to handle those constant changes in ability, especially because it is nearly impossible for the employees to keep track of all their colleagues' skills over time.

1.1.5 Technologies

The efficiency of EFSs is closely tied to the different technologies that are being used. Therefore, the following components and technologies are of interest for the EFS:

- **Reliable data:** Data quality is one of the most important factors for the success of an EFS as it is the basis for the search algorithm. Some of the more popular data sources of commercial tools are Self-declared data, Documents and Databases (M. T. Maybury 2006, p. 18)
- **Search algorithm:** The search algorithm is the core of the EFS. It has to be able to handle the data and provide the user with the most relevant results. Here Keyword search, and Boolean search are the most common methods with the Natural Language Search which utilizes NLP being on the third place, though since the release of the journal article by M. T. Maybury (ibid.) (ibid., p. 18), NLP has gained a lot of popularity especially through the rise of AI and Machine Learning with applications in Chatbots, Voice Assistants and Sentiment analysis (Administrator 2023).
- **UI:** The UI is the interface between the user and the EFS. It has to be intuitive and easy to use in order for the user to actually use the EFS. The UI of an EFS most commonly consists of different components like a search bar, a result overview and detail pages for each result based on the review of big EFS like LinkedIn (*LinkedIn* 2024), Research Gate (*ResearchGate | Find and share research* 2024) or Expertise Finder (*Expertise Finder | Expert Systems, Online Directory* 2024). Regarding the results overview, a list of experts seems to be the most common approach, with related documents and related concepts also being provided in some cases. (M. T. Maybury 2006, p. 18)
- **Administration and Maintenance:** The EFS has to be maintained and updated regularly in order to keep up with the changes in the organization. This includes an Admin-Panel for the administrators to manage the data, as well as a feedback system for the users to report changes and errors in the data.

In addition, FastAPI, React with Scale and Google Firebase are used in this project. FastAPI is an easy-to-use web framework for building python-based Application Programming Interfaces (APIs) (*FastAPI* 2024) and is used as the projects' backend with the main purpose of handling the AI based search. React is the frontend framework of choice. It is the most popular web framework among professional developers according to *Technology | 2024 Stack Overflow Developer Survey* (2024) with 41.6% of the 38132 responses. It offers great documentation and reusable components. In combination with React, scale is the component library for the system. It is an open-source library providing production-ready UI components with brand compliance and accessibility in mind (*FAQ - Page Storybook* 2024). Lastly, Google Firebase is used as the database for the EFS. Firebase offers a variety of different services both for development and production like the database, authentication, hosting and analytics. Firebase was chosen, for its ease of implementation and great documentation.

1.2 Research Question

The main research question of the thesis is:
How can an Expert Finder system for corporate structures be designed and implemented?
In order to answer this question, is broken down as follows:

- *What are the key functionalities such an expert finder should possess?*
- *How can the user interface be designed to enable intuitive usage?*
- *Which AI supported search method is best suited for the expert finder with respect to accuracy, scalability and ease of integration?*
- *How can a workflow system be implemented to provide the user with concrete recommendations for solving specific problems?*

1.3 Problem Statement

In large companies and organizations, identifying and finding experts is a common challenge. As a company grows, so does the complexity of its internal organization, making it increasingly difficult for the companies as well as their employees to keep an overview over the existing experts and contact persons. Factors like a rise in mobile work and agile teams only act as a catalyst towards this problem. This gets problematic, as it is crucial for employees to be able to find the right contact persons or experts to help them fulfill tasks more efficiently. When not addressed, this problem results in different consequences, examples being loss of time and resources due to unproductivity, duplication of work as well as missed opportunities for collaboration.

Traditional systems address this problem by enabling users to search for experts and providing the contact data of the resulting experts. Those systems often times are focused primarily on the transfer of knowledge and don't emphasize enough on task completion. This often results in users having to take the extra step to contact the found experts outside the EFS, ending up in unnecessary effort both for the users and experts. The users, as they have to contact the experts themselves, and the experts, as users often times fail on providing all the necessary information outright. This challenge has been informally noted by several colleagues in the company where I work, who mentioned that experts often have to ask for follow-up information in order to gather missing details, leading to further time-inefficiencies. In addition, systems most of the time are very costly to develop and maintain as can be seen in M. T. Maybury (2006, pp. 19–36), with recent advancements in Large Language Model (LLM) and NLP leaving room for improvement to realize a better user experience.

1.4 Structure of the Thesis

This thesis is structured in seven chapters to provide the reader with a comprehensive understanding of the topic and the research conducted. The first chapter introduces the topic in highlighting the background of the thesis as well as the research question. It also defines the scope and delimitations of the thesis. The second chapter outlines previous research conducted on the topic of EFSs and provides the fundamentals of the technologies utilized in this thesis. Additionally, it identifies the research gap that this thesis addresses. In the third chapter the methods used for research and development are highlighted to provide the reader with an understanding of the research process. The fourth chapter presents the implementation of the EFS and discusses technical details of the implemented system. The fifth chapter evaluates the proposed system in establishing evaluation criteria and analyzing feedback and performance. It also tries to address the strengths and weaknesses of the system. In the sixth chapter, the results of the previous chapters are interpreted and discussed. Lastly, the seventh chapter concludes the work by summarizing the findings, pointing out potential optimizations and proposing future work.

1.5 Scope and Delimitations

- **Scope:**
  The study focuses on designing and developing a prototype aimed at providing the user with recommendations of experts and workflows fit to their needs. The scope

includes four main components. Firstly, an AI supported search aimed at providing the user with the most relevant results for a given query. Secondly, a database of experts, workflows and topics managed by an administration panel, for the administrators to create update and delete data. At third, a workflow builder that enables administrators to create workflows aimed at directly solving the users' needs without the need of additional contact to the expert, and lastly a user-friendly UI that enables the user to easily navigate the system, browse over experts and transparently communicates the quality of the results. The prototype is designed to initially be used by a small, selected group of people, with the potential to be scaled up over time. The concept is especially intended for larger structures, where seamless communication and collaboration is very important but often hard to achieve.

- **Delimitations:**
  Regarding AI the study specifically inspects how it can enhance the search for experts and workflows in an expert and workflow database, and which AI technology fits the best for the given use case. This does not include AI supported expert identification or validation and instead provides administrators with an option to add experts and workflows themselves. Also, the workflow builder will be limited to just the most important features to prove the concept. While privacy and data protection will be considered to a degree, the primary focus of the study is on functionality and efficiency of the prototype.

2 State of Research

The following chapter offers an overview on the state-of-the-art research on EFSs as well as the concepts and technologies related to the topic and used in the prototyping and implementation, to give a basic understanding on the topic. This is particularly important in order to take previous findings into account for this work and to highlight research gaps. The chapter is split into 5 sub-topics which focus on EFSs as such, the basics of the underlying technologies, UI design of EFSs, the integration of workflows in EFSs and finding research gaps in this field of research.

2.1 Overview of Expert Finding System

2.1.1 Historical Development

The Topic of EFSs has now been researched for around 35 years. One of the first studies in this field was conducted by Ackerman and Malone (1990) in the conference paper ‘Answer Garden’. Answer Garden was designed as a tool, allowing organizations to implement a database of Frequently Asked Questions (FAQ) and corresponding answers, in order to better utilize the knowledge of their knowledgeable employees. The tool was aimed at user-groups that often times had to repeatedly answer the same questions, with occasionally new questions not answered before. The system on the one hand allowed users to explore the database of questions and answers with a branching network, a component functioning like a decision tree by guiding a user through a series of choices, with the goal of providing the user with the fitting answer. On the other hand the system will, in case no fitting answer is available, direct the users´question to a matching expert, whose answer is then returned to both the user and the database, so other users won’t have to consult an expert for this specific question no more (ibid., pp. 1, 3). Regarding the identification of experts, systems of this time relied on the manual input of experts, which compared to current systems is much more time-consuming. A few years later, the journal article ‘Discovering shared interests using graph analysis’ by Schwartz and Wood (1993) was published. In comparison to Answer Garden, the system was not specifically designed to answer specific questions but rather to find people with specific interests or expertise. Regarding the aspect of people finding, in contrast to the typical approach where a directory from explicitly registered data was used for locating people, a system displaying shared interests and expertise of different people was proposed (ibid., pp. 1–2). To achieve this, e-mail communication was analyzed via graph analysis (ibid., p. 9). The approach was chosen to not only solve the "white page" problem, describing the issue of locating a particular user, but also address the "yellow page" problem, which describes the issue of finding a user with a specific interest or expertise (Schwartz and Wood 1993, p. 1).

2.1.2 Different Types of Expert Finding System

2.1.2.1 Knowledge-Transfer-Oriented Expert Finding System

2.1.2.2 Task-Oriented Expert Finding System

2.2 Technological Foundations

2.2.1 Large Language Models, Natural Language Processing, and Machine Learning in Expert Finding Systems

2.2.1.1 Large Language Models

2.2.1.2 Natural Language Processing

2.2.1.3 Machine Learning

2.2.2 Frontend and Backend Technologies used

2.2.2.1 React (Frontend)

2.2.2.2 FastAPI (Backend)

2.2.2.3 Role of Application Programming Interfaces in Expert Finding Systems

2.2.3 Data Sources and Data Quality

2.2.3.1 Reliability of Data

2.2.3.2 Common Data Sources in Expert Finding Systems

2.2.4 Self-Hosted vs. Third-Party Solutions

2.2.4.1 Self-Hosted Solutions

2.2.4.2 Third-Party Solutions

2.3 Integration of Workflows in Expert Finding Systems

2.3.1 Task-Oriented Design in Expert Finding Systems

2.3.2 Automation and AI integration

2.3.3 Challenges and Solutions in Workflow Integration

3 Methodology

3.1 System Design

3.2 UI Design

3.3 Workflow Integration

3.4 Prototype Development

4 Implementation

4.1 Architecture and Design

4.2 Technical Challenges and Solutions

4.3 Testing and Debugging

5 Evaluation

5.1 Evaluation Criteria

5.2 User Testing

5.3 Strengths and Weaknesses

6 Conclusion and Outlook

6.1 Conclusion

6.2 Outlook

Bibliography

Ackerman, M. S. and T. W. Malone (1990). 'Answer Garden: a tool for growing organizational memory'. In: *Proceedings of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems*. COIS90: Conference on Office Infor. Systems. Cambridge Massachusetts USA: ACM, 1990-03, pp. 31–39. ISBN: 978-0-89791-358-4. URL: https://dl.acm.org/doi/10.1145/91474.91485 (visited on 2024-09-09).

Administrator (2023). *The Role of Natural Language Processing in AI*. University of San Diego Online Degrees. 2023-03-28. URL: https://onlinedegrees.sandiego.edu/natural-language-processing-overview/ (visited on 2024-08-10).

Becerra-Fernandez, I (2000). 'The role of artificial intelligence technologies in the implementation of People-Finder knowledge management systems'. In: *Knowledge-Based Systems* 13.5 (2000-10), pp. 315–320. ISSN: 09507051. URL: https://linkinghub.elsevier.com/retrieve/pii/S0950705100000915 (visited on 2024-08-09).

Becerra-Fernandez, Irma (1999). 'Searchable Answer Generating Environment (SAGE): A Knowledge Management System for Searching for Experts in Florida'. In: (1999).

Becerra-Fernandez, Irma (2000). 'Facilitating the Online Search of Experts at NASA using Expert Seeker People-Finder'. In: (2000).

Expertise Finder | Expert Systems, Online Directory (2024). URL: https://expertisefinder.com/,%20https://expertisefinder.com/ (visited on 2024-08-10).

FAQ - Page Storybook (2024). URL: https://telekom.github.io/scale/?path=/docs/faq–page (visited on 2024-09-08).

FastAPI (2024). URL: https://fastapi.tiangolo.com/ (visited on 2024-09-08).

Husain, Omayma, Naomie Salim, Rose Alinda Alias, Samah Abdelsalam, and Alzubair Hassan (2019). 'Expert Finding Systems: A Systematic Review'. In: *Applied Sciences* 9.20 (2019-01). Number: 20 Publisher: Multidisciplinary Digital Publishing Institute, p. 4250. ISSN: 2076-3417. URL: https://www.mdpi.com/2076-3417/9/20/4250 (visited on 2024-08-09).

Inc, Gallup (2024). *State of the American Workplace*. Gallup.com. URL: https://www.gallup.com/workplace/238085/state-american-workplace-report-2017.aspx (visited on 2024-08-08).

LinkedIn (2024). URL: https://www.linkedin.com/ (visited on 2024-08-10).

Mattox, David, Mark Maybury, and Daryl Morey (1999). *Enterprise expert and knowledge discovery*. Journal Abbreviation: Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany Pages: 307 Publication Title: Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany. 1999-01-01. 303 pp.

Maybury, Mark T (2006). 'Expert Finding Systems'. In: (2006-09).

Organ, Christine (2023). *7 Organizational Structure Types (With Examples)*. Forbes Advisor. Section: Business. 2023-04-07. URL: https://www.forbes.com/advisor/business/organizational-structure/ (visited on 2024-08-07).

ResearchGate | Find and share research (2024). ResearchGate. URL: https://www.researchgate.net/ (visited on 2024-08-10).

Schwartz, Michael F. and David C. M. Wood (1993). 'Discovering shared interests using graph analysis'. In: *Communications of the ACM* 36.8 (1993-08), pp. 78–89. ISSN: 0001-

0782, 1557-7317. URL: https://dl.acm.org/doi/10.1145/163381.163402 (visited on 2024-09-10).

Stankovic, Milan, Claudia Wagner, Jelena Jovanovic, and Philippe Laublet (2010). *Looking for Experts? What can Linked Data do for You?* Vol. 628. Journal Abbreviation: CEUR Workshop Proceedings Publication Title: CEUR Workshop Proceedings. 2010-01-01.

Technology | 2024 Stack Overflow Developer Survey (2024). URL: https://survey.stackoverflow.co/2024/technology#most-popular-technologies-webframe-prof (visited on 2024-09-08).

What is a Functional Structure in an Organization? (2024). *What is a Functional Structure in an Organization? | Miro*. https://miro.com/. URL: https://miro.com/organizational-chart/what-is-a-functional-structure/ (visited on 2024-08-07).

What is Matrix Organization? (2024). *What is Matrix Organization? Definition of Matrix Organization, Matrix Organization Meaning*. The Economic Times. URL: https://economictimes.indiatimes.com/definition/matrix-organization (visited on 2024-08-08).

Reference of AI Tools

I hereby declare that I have used the following AI tools for this thesis:

- DeepL for translating parts of the thesis from German to English
- GitHub Copilot for code suggestions and completions for the LaTeX code
- OpenAI ChatGPT for supporting the brainstorming ideas for the thesis

Declaration in lieu of oath

I hereby declare that I produced the submitted paper with no assistance from any other party and without the use of any unauthorized aids and, in particular, that I have marked as quotations all passages which are reproduced verbatim or near-verbatim from publications. Also, I declare that the submitted print version of this thesis is identical with its digital version. Further, I declare that this thesis has never been submitted before to any examination board in either its present form or in any other similar version. I herewith agree that this thesis may be published. I herewith consent that this thesis may be uploaded to the server of external contractors for the purpose of submitting it to the contractors' plagiarism detection systems. Uploading this thesis for the purpose of submitting it to plagiarism detection systems is not a form of publication.

Nürnberg, 10.9.2024
(Location, Date)

(handwritten signature)