Business User-oriented Recommender System of Data

Pinon, Sarah

DOI:
10.1007/978-3-031-33080-3_44

Publication date:
2023

Document Version
Peer reviewed version

Business User-oriented Recommender System of Data

Sarah Pinon[0000−0001−5392−1020]

NaDI, Namur Digital Institute, University of Namur, Belgium
sarah.pinon@unamur.be

Abstract. Companies nowadays are increasingly dependent on data. In an environment that is more dynamic than ever, they are looking for tools to leverage those data and obtain valuable information in a rapid and flexible way. One way to achieve this is by using Data-Driven Decision Support Systems (Data-Driven DSS). In this project, I focus on one such type of DSS, namely the Self-Service Business Intelligence (SSBI) Systems. These systems are designed specifically to avoid the involvement of the IT department when creating business reports by empowering businesspeople in the production of their own reports, thereby reducing the time-to-release of a given report and improving the responsiveness of companies. Business decision-makers, when developing their own reports, however face barriers. These challenges are related to the current self-service features that are not sufficiently adapted to their business needs and their lack of technical knowledge. The objective of my project is to build a framework based on Artificial Intelligence (AI) techniques such as Natural Language Processing techniques, Semantic and Recommender Systems to solve one of the main challenges faced by businesspeople, namely: the data picking within technical and large current databases. These AI systems offer a number of benefits that are strongly linked to the problems encountered by business users in the data picking process. This paper introduces the three main research questions of my thesis and positions them in the current literature. It then elaborates on the different theoretical, methodological and empirical contributions I plan to advance as part of my project.

Keywords: Self-Service Business Intelligence · Business User · Data · Recommender System · Semantic Systems · Natural Language Processing.

1 Context

The current success of companies depends on their ability to use the available data [6]. In fact, companies that make decisions based on data are more productive than companies that rely on intuition [21]. Extracting useful knowledge from data to solve business problems is becoming a priority for companies in almost every industry [21]. However, this task is increasingly complex as the volume

of data continues to grow. Companies are therefore investing in Decision Support Systems (DSS) to manage this amount of information [6]. These systems are nowadays a business necessity but also an opportunity to gain a significant competitive advantage [20].

A Business Intelligence (BI) is one such tool that assists decision makers in their process by transforming data into information and knowledge[29]. Indeed, BI systems collect data, store it and produce knowledge with analytical tools to present complex information to decision makers[16]. Several actors are involved in the BI process. These actors can be divided into two groups. The first are IT professionals or experienced BI users. They have the technical skills to efficiently set up and run BI systems. The second group consists of decision-makers, called Business Users. They do not have technical skills and use BI tools to support their decision making [2, 14].

Over the last few years, BI had to adapt to a new type of requirement from companies that face an increasingly dynamic environment, and which as a result need to be more and more reactive in their decision-making[14, 31]. To meet this new requirement, a new version of BI tools is being developed, called Self-Service Business Intelligence (SSBI)[13, 14].

SSBI systems aim to provide reports and dashboards to decision makers much faster. To achieve this, these systems allow decision makers to build reports and dashboards themselves. In this way, all the necessary interactions between IT experts and business users to define the report requirements are eliminated and a considerable amount of time is saved [30, 31].

SSBI tools are increasingly used in companies[31]. Therefore, there is a growing interest in their efficiency in the literature. Nowadays, it turns out that SSBI tools are not as easy to use as they promise [14]. This issue comes from the many challenges faced by Business Users throughout the BI process, from the access to data to the interpretation of reports produced [13, 14].

In my research project, I focus on one of the challenges faced by Business Users: finding relevant data fields within the current large and technical databases. There is a lot of evidence showing the difficulties business people face when dealing with databases, such as exploring them, locating specific data or identifying data fields for a particular need [25, 13, 14, 2]. These challenges imply three negative effects. First, the adoption of SSBI tools is hindered, even though it is critical in today's dynamic business environment [14]. Second, these obstacles imply an under-exploitation of the available data, which remains not understood by the user. Third, there is a high risk of wrong data selection, which can compromise the quality of decision outputs [13, 14].

This paper is organized as follows: Section 2 explains the research objectives of the project by developing the state of the problem and the research questions. Section 3 develops the research approach by explaining for each research question the objective, methods used, related work, current achievement and preliminary results. Section 4 concludes the paper.

2 Research objectives

2.1 Problem statement

The challenges faced by Business Users when dealing with current databases are mainly related to two of their characteristics:

1. **The amount of data fields**: In the era of Big Data, the volume of available data is continuously increasing [28]. This quantitative growth comes from technological advances in digital sensors, communication, computation, and storage[8]. While collecting data is becoming easier, identifying relevant information within that data is more and more complex. Indeed, several data fields do not provide any interesting information or do not concern all the Business Users of the company [14]. This task of identification requires special skills and technologically-adapted tools [21, 28]. Business Users have limited IT skills and therefore have difficulties to perform the task properly [14].
2. **Technical jargon**: Databases frequently use highly technical jargon. The terminologies employed are not necessarily familiar to Business Users. Indeed, the labels of the data can, for example, correspond to names of codes or can have several definitions according to the type of Business User [14]. This jargon implies a lack of understanding of the data meaning for the users [24].

In addition to the complexity of today's databases, Business Users do not always have a clear vision of what data they actually need [14]. There are many situations where the decision maker does not know from which angle to analyze a company: Is it the marketing actions that need to be reviewed? Is it the performance of the employees that needs to be checked? Is the cost of products too high? The volume and technicality of today's databases will further obscure their fuzzy vision of what they are looking for.

The ultimate goal of my project is to help Business Users with these challenges by providing recommendations about which data from a large data repository should be used by the decision maker, based on question submits with his/her own words and without any knowledge of the underlying naming of the database, tables or columns. Moreover, when producing these recommendations, I take care to adopt a broad perspective on all possible data, opening the eyes of Business Users to the field of possibilities. In this way, I optimize the value generated by the data for decision makers.

To achieve this, I propose a solution mobilizing different Artificial Intelligence (AI) techniques, namely: Natural Language Processing (NLP) techniques, Semantic Systems (e.i. Ontologies) and Recommender Systems (RSs). These different techniques offer indeed many advantages to help users to face these challenges.

NLP techniques aim to understand and manipulate the natural languages of human beings. These techniques are, for example, used within user interfaces to translate requests expressed in natural language into requests expressed in

language understandable by the system [5]. These techniques are more and more exploited within SSBI tools in order to allow Business Users to explore their data in a more natural way and without requiring the help of IT experts [18, 1].

Ontologies are semantic modeling systems. They offer a formal semantic representation of the system's structure[9, 17]. By using Ontologies within SSBI, the business semantics used by the Business Users can be integrated to, for example, define the different data of the company's Data Warehouse (DWH) [23, 4]. This business jargon simplifies and makes more intuitive the interaction between the Business User and the system. One can indeed speak of business-oriented interaction rather than technology-centered interaction[4].

Regarding the RS, its objective is to generate meaningful recommendations to users for items or products that might interest them [15]. In the context of SSBI, the RS can allow to detect the Business Users's interests and thus to characterize the type of interaction desired with the SSBI tool [7]. These systems can then guide and support the Business User in his/her process of transforming data into information [26].

2.2 Research Questions

The previous large objective brings me to the definition of a number of research questions:

- **RQ1**: How to help Business Users to pick relevant data within a voluminous and technical DWH through the Business Glossary and a Semantic RS ?
- **RQ2**: What are the barriers to adoption of business decision makers using SSBI, and how does our implemented Data RS help mitigate those barriers ?
- **RQ3**: How can we account for the user profile in the Data RS to offer personalized data suggestions ?

3 Research Approach

This PhD research is currently composed of three ongoing studies corresponding to the three research questions. Each study has its own methodology and contribution:

1. Data RS for SSBI tools;
2. Empirical study on Data RS;
3. Business User-centric RS of Data.

The three studies together follow the methodology called Design Science Research methodology[19, 10]. This methodology aims to create new innovative artifacts that solve real-world problems. This approach develops and evaluates new solutions[19, 10]. My research project follows this approach through these three studies. Indeed, the first research consists in developing my solution answering a problem identified in the literature. This step corresponds to the rigor

cycle of the methodology. Then, the empirical study of this research consisting in demonstrating the solution relevance for the companies and their employees refers to the relevance cycle of the methodology. Finally, the last study of the project corresponds to an improvement of the system involving all the cycles of the methodology, namely the existing literature and a field study of the users of the existing tools.

3.1 Study 1 - Data RS for SSBI tools

Objective. The first step of my research project is to support Business Users to pick the right data, among technical and voluminous databases, based on a request expressed in semi-natural language. Concretely, my data picking support system will allow these users to express their data needs with their own business words without necessarily referring to the technical jargon of the database. Compared to what is proposed in the literature, my system will exploit the Business Glossary associated with the database, develop a Data RS and offer a great transparency and flexibility.

Methods. To achieve the objective of this paper, I exploit RSs. These systems help people to cope with information overload [22] which is completely aligned with the problem my system is facing, namely data overload. Moreover, my RS is based on an ontological representation of the DWH schema integrating the associated business semantics via the Business Glossary. Ontology-based RSs allow to deal with most of the conventional RS problems such as cold-start problem, rating sparsity and overspecialization [27]. As for the Business Glossary, it allows to add the business context to the available data [12]. To provide the user with data recommendations, my system exploits a NLP technique and a set of business rules to link the terms used by the user in his/her request to the numerous technical data fields of the DWH.

Related Work. Systems to make databases more easily usable by Business Users have already been developed in the literature. Some papers develop, indeed, Natural Language Interface for Databases (NLIDB). These systems transform data queries expressed in natural or semi-natural language into complex languages such as SQL [18, 1]. Other works exploit Semantic Systems to add a semantic layer to databases [25, 23]. To the best of my knowledge, no Data RS has been already developed.

Current Achievement and Preliminary Results. This first step of my research project is already implemented and the associated paper is being written.

3.2 Study 2 - Empirical study on Data RS

Objective. To address the second research question of my project: "What are the barriers to adoption of business decision makers using SSBI, and how does

our implemented Data RS help mitigate those barriers ?", I will study my problem of interest in a real-world setting and develop my RS developed in the first study within several industries in order to identify its added value. The companies used as use cases will be a large consortium of hospitals thanks to a collaboration with the research center CETIC ("Centre d'Excellence en Technologies de l'Information et de la Communication") and the international company TotalEnergies.

Methods. First, a qualitative study will be conducted with the business decision-makers of the two collaborators mentioned above in order to identify and confirm the challenges they face with large and technical databases. Second, after developing our RS on the data of our two partners, we will do another qualitative study to demonstrate the added value of our system and identify its limitations.

Related Work. To the best of my knowledge, no study has been done specifically on the limitations encountered by Business Users in relation to current databases. Moreover, as this study serves as a use case for the system we developed in the first article, the contribution is completely new to the literature.

Current Achievement and Preliminary Results. This phase of the project has not yet been implemented. Only collaborative initiatives with businesses have been launched.

3.3 Study 3 - Business User-centric Data RS

Objective. In this study, I will try to answer the third question of my research: "How can we account for the user profile in the Data RS to offer personalized data suggestions ?". This paper aims to offer an advanced version of our tool developed in the first study. Indeed, I want to integrate the user's profile (e.g. his/her data preferences, historical interactions) as well as additional functionalities (e.g. feedback on recommendations). These different features will allow me to refine my recommendations.

Methods. To improve the RS, I will exploit the weighting of ontologies to, for example, represent the user's historical interactions with certain data on the DWH ontology. The weights are then exploited by the RS to recommend the most relevant fields, as is done by the RS based on the case-based approach [3]. In addition to the ontology weighting, this study will identify the different data needs of different Business User profiles in order to help the RS. To do this, a qualitative study will be conducted with different Business Users from different companies. I will identify their data needs and analyze the differences between the different users according to their profile. This study will be inspired by what has been done by [11] for the audience of visualization technologies.

Related Work. As previously explained, no Data RS has been, to my knowledge, developed in the current literature. Consequently, no Data RS considering

the user's profile and my additional features exists. Concerning the user profile, some studies have been conducted to characterize, not the Business Users in the context of Data Picking within DWH, but, for example, the audience of visualization technologies [11].

Current Achievement and Preliminary Results. During the first months of research, an exploration of the different existing RSs was carried out. This allowed me to understand how they work, their advantages and disadvantages. I also analyzed different studies studying the categorization of BI users in terms of visualization tools for example.

4 Conclusion

This PhD research proposes a complete system to support business decision-makers in the process of data picking within large and technical databases. The proposed system is completely adapted to the business world by considering the business semantic and the user's profile. I suggest integrating AI techniques and in particular, NLP techniques, Semantic Systems and RSs, to achieve this goal. The research project will be divided into three studies. These studies will provide methodological contributions to the existing literature to address the problem identified and developed in this paper. Fundamental contributions will also be presented in these different studies by the definitions of several ontologies and the integration methods of these knowledge models.

Acknowledgements I would like to thank the supervisors of this project, Dr.Corentin Burnay and Dr. Isabelle Linden, for their review and support on this paper. Moreover, this research was financed by the Walloon Region in the scope of the ARIAC project.

References

1. Affolter, K., Stockinger, K., Bernstein, A.: A comparative survey of recent natural language interfaces for databases. The VLDB Journal **28**(5), 793–819 (2019)
2. Alpar, P., Schulz, M.: Self-service business intelligence. Business & Information Systems Engineering **58**(2), 151–155 (2016)
3. Bridge, D., Göker, M.H., McGinty, L., Smyth, B.: Case-based recommender systems. The Knowledge Engineering Review **20**(3), 315–320 (2005)
4. Cao, L., Zhang, C., Liu, J.: Ontology-based integration of business intelligence. Web Intelligence and Agent Systems: An International Journal **4**(3), 313–325 (2006)
5. Chowdhary, K.: Natural language processing. Fundamentals of artificial intelligence pp. 603–649 (2020)
6. Cody, W.F., Kreulen, J.T., Krishna, V., Spangler, W.S.: The integration of business intelligence and knowledge management. IBM systems journal **41**(4), 697–713 (2002)

7. Drushku, K., Aligon, J., Labroche, N., Marcel, P., Peralta, V., Dumant, B.: User interests clustering in business intelligence interactions. In: International Conference on Advanced Information Systems Engineering. pp. 144–158. Springer (2017)
8. Emani, C.K., Cullot, N., Nicolle, C.: Understandable big data: a survey. Computer science review **17**, 70–81 (2015)
9. Guarino, N., Oberle, D., Staab, S.: What is an ontology? In: Handbook on ontologies, pp. 1–17. Springer (2009)
10. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. Management Information Systems Quarterly **28**(1), 6 (2008)
11. Kerren, A., Stasko, J., Fekete, J.D., North, C.: Information visualization: human-centered issues and perspectives, vol. 4950. Springer (2008)
12. Kotsis, G., Tjoa, A.M., Khalil, I., Moser, B., Mashkoor, A., Sametinger, J., Fensel, A., Martinez-Gil, J., Fischer, L., Czech, G., et al.: Database and Expert Systems Applications-DEXA 2021 Workshops: BIOKDD, IWCFS, MLKgraphs, AICARES, ProTime, AISys 2021, Virtual Event, September 27–30, 2021, Proceedings. Springer Nature (2021)
13. Lennerholt, C., van Laere, J., Söderström, E.: Implementation challenges of self service business intelligence: A literature review. In: 51st Hawaii International Conference on System Sciences, Hilton Waikoloa Village, Hawaii, USA, January 3-6, 2018. vol. 51, pp. 5055–5063. IEEE Computer Society (2018)
14. Lennerholt, C., Van Laere, J., Söderström, E.: User-related challenges of self-service business intelligence. Information Systems Management **38**(4), 309–323 (2021)
15. Melville, P., Sindhwani, V.: Recommender systems. Encyclopedia of machine learning **1**, 829–838 (2010)
16. Negash, S., Gray, P.: Business intelligence. In: Handbook on decision support systems 2, pp. 175–193. Springer (2008)
17. Obrst, L.: Ontologies for semantically interoperable systems. In: Proceedings of the twelfth international conference on Information and knowledge management. pp. 366–369 (2003)
18. Őzcan, F., Quamar, A., Sen, J., Lei, C., Efthymiou, V.: State of the art and open challenges in natural language interfaces to data. In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. pp. 2629–2636 (2020)
19. Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J.: Design science research process: A model for producing and presenting information systems research. arXiv preprint arXiv:2006.02763 (2020)
20. Power, D.J.: Decision support systems: concepts and resources for managers. Greenwood Publishing Group (2002)
21. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. Big data **1**(1), 51–59 (2013)
22. Resnick, P., Varian, H.R.: Recommender systems. Communications of the ACM **40**(3), 56–58 (1997)
23. Sell, D., da Silva, D.C., Beppler, F.D., Napoli, M., Ghisi, F.B., Pacheco, R.C., Todesco, J.L.: Sbi: a semantic framework to support business intelligence. In: Proceedings of the first international workshop on ontology-supported business intelligence. pp. 1–11 (2008)
24. Smuts, M., Scholtz, B., Calitz, A.: Design guidelines for business intelligence tools for novice users. In: Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists. pp. 1–15 (2015)

25. Spahn, M., Kleb, J., Grimm, S., Scheidl, S.: Supporting business intelligence by providing ontology-based end-user information self-service. In: Proceedings of the First international Workshop on ontology-Supported Business intelligence. pp. 1–12 (2008)
26. Sulaiman, S., Gómez, J.M.: Recommendation-based business intelligence architecture to empower self service business users. In: Multikonferenz Wirtschaftsinformatik. pp. 1–12 (2018)
27. Tarus, J.K., Niu, Z., Mustafa, G.: Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial intelligence review **50**(1), 21–48 (2018)
28. Tsai, C.W., Lai, C.F., Chao, H.C., Vasilakos, A.V.: Big data analytics: a survey. Journal of Big data **2**(1), 1–32 (2015)
29. Vercellis, C.: Business intelligence: data mining and optimization for decision making. John Wiley & Sons (2011)
30. Weiler, S., Matt, C., Hess, T.: Understanding user uncertainty during the implementation of self-service business intelligence: A thematic analysis. In: Hawaii International Conference on System Sciences (HICSS) (2019)
31. Yu, E., Lapouchnian, A., Deng, S.: Adapting to uncertain and evolving enterprise requirements: The case of business-driven business intelligence. In: IEEE 7th International Conference on Research Challenges in Information Science (RCIS). pp. 1–12. IEEE (2013)