Article

Conversation Concepts: Understanding Topics and Building Taxonomies for Financial Services

John P. McCrae [1,*], **Pranab Mohanty** [2], **Siddharth Narayanan** [2], **Bianca Pereira** [1], **Paul Buitelaar** [1], **Saurav Karmakar** [1] and **Rajdeep Sarkar** [1]

1   Insight SFI Research Centre for Data Analytics, Data Science Institute, NUI Galway, Galway H91 A06C, Ireland; bianca.pereira@insight-centre.org (B.P.); paul.buitelaar@nuigalway.ie (P.B.); saurav.karmakar@insight-centre.org (S.K.); rajdeep.sarkar@insight-centre.org (R.S.)
2   FMR LLC, Boston, MA 02110, USA; pranab.mohanty@fmr.com (P.M.); siddharth.narayanan@fmr.com (S.N.)
*   Correspondence: john@mccr.ae

Abstract: Knowledge graphs are proving to be an increasingly important part of modern enterprises, and new applications of such enterprise knowledge graphs are still being found. In this paper, we report on the experience with the use of an automatic knowledge graph system called Saffron in the context of a large financial enterprise and show how this has found applications within this enterprise as part of the "Conversation Concepts Artificial Intelligence" tool. In particular, we analyse the use cases for knowledge graphs within this enterprise, and this led us to a new extension to the knowledge graph system. We present the results of these adaptations, including the introduction of a semi-supervised taxonomy extraction system, which includes analysts in-the-loop. Further, we extend the kinds of relations extracted by the system and show how the use of the BERTand ELMomodels can produce high-quality results. Thus, we show how this tool can help realize a smart enterprise and how requirements in the financial industry can be realised by state-of-the-art natural language processing technologies.

Keywords: knowledge graphs; FinTech; taxonomies; financial services; natural language processing; relation extraction; term extraction

1. Introduction

Enterprise knowledge graphs [1] are a key tool that can enable businesses to create smart artificial-intelligence-driven applications that can bring real business value. However, two principle issues often face businesses who wish to use an enterprise knowledge graph: firstly, the construction of a knowledge graph is neither inexpensive nor straightforward, and secondly, the applications of the enterprise knowledge graph are not clear. In this paper, we present our solution to and experiences with this problem, drawn from a collaborative project between an academic institution, the National University of Ireland (NUI) Galway, and an enterprise, FMR LLC.

This collaboration builds firstly on the work of the Insight SFIResearch Centre for Data Analytics at NUI Galway's experience with the creation of knowledge graphs and, in particular, the system called Saffron [2]. This system is designed to create a taxonomy automatically from a large text corpus and works by first identifying all the terms using syntactic and corpus frequency information. Then, relationships between these terms are found, and these terms are organized into a taxonomy structure.

These tools are planned to be integrated into a new application called Conversation Concepts Artificial Intelligence (CCAI), which is used by FMR LLC to drive business insights. The objectives of this application were distilled into use cases that could be realised by state-of-the-art natural language processing technologies. In particular, there was a strong desire to include analysts within the taxonomy construction process. This led to the development of a new semi-automatic methodology for taxonomy development, which

allows the results to be updated and then dynamically updates the results of the automatic extraction process to take into account these changes. Secondly, the need for intent classification within a virtual agent system led to a use case of extending the taxonomy extraction to extract a wide range of relations including synonymy and meronymy. As such, we are just starting to understand the ways that enterprise knowledge graphs can produce real business value, and as such, we are still discovering new applications, such as customer satisfaction analysis, and how these can be handled by NLP applications.

The rest of this paper is structured as follows: firstly, we briefly describe the state-of-the-art in enterprise knowledge graphs in Section 2. We then describe the natural language processing research, which goes into the development of Saffron and the CCAI application in Section 4. Then, we describe the CCAI applications and some of the business use cases, as well as some of the future envisioned applications in Section 3. Finally, we discuss how these results could be applied within other enterprises and how the goal of a smart enterprise can be attained in Section 5 before providing some concluding remarks Section 6

2. Related Work

Enterprise knowledge graphs [1] originally sprung from the ideas of semantic networks as proposed by Quillian [3] and have increasingly become an important part of large enterprises. These knowledge graphs are founded on the work of the Semantic Web [4] and linked data [5] and make use of formalisms such as RDF [6] and OWL [7]. Interest in such graphs has been further sparked by their adoption by large corporations such as Google [8]. This has led to knowledge graphs being seen as a key part of Enterprise 2.0, and large platforms such as miKrow [9] have been developed to simplify the creation of knowledge graphs and improve user interaction within enterprises.

Denaux et al. [10] proposed a reference architecture for enterprise knowledge graphs consisting of three main sections:

Acquisition This is concerned with the development of the knowledge graph and includes the construction and learning of a knowledge graph from data.
Storage A knowledge graph is often very large, and so, efficient query and access to the knowledge graph can be a key technical challenge
Consumption In order to unlock the value of a knowledge graph, it is necessary to develop friendly user-facing applications that can deliver genuine business value.

In this paper, our focus is primarily on the first and third part of the architectures.

The acquisition of knowledge graphs is primarily based on the life-cycle of the knowledge graph [11], which involves the extraction, revision and evolution of the knowledge graph in order to ensure that the knowledge graph is of high quality. The process of extracting a knowledge graph is essentially the same as that of ontology learning [12] and consists of three main stages: firstly, the relevant entities or terms must be identified, in a process called automatic term recognition, then the relations between these terms must be constructed, and finally, these must be organized into a single coherent knowledge graph. For the first task of extracting terms, a number of heuristic methods have been developed based on the frequency of terms and syntactic patterns [2,13,14]. Relation extraction approaches were initially based on the discovery of patterns that are indicative of relations, such as the patterns proposed by Hearst [15]; however, recently, new distributional methods have become popular for extracting relations [16], which rely on word embeddings to predict relations between entities, and authors have explored combining pattern-based and distributional methods [17] or extending the results to multiple relations [18]. Taxonomy construction extends the task of term and relation extraction to a holistic approach, and initial results have shown that this is a very hard task [19]; until recently, very few approaches have improved on the state-of-the-art [2,20].

For the application of knowledge graphs, question answering and chatbot systems have been identified as key targets for the application of knowledge graphs [21]. Some early systems approached this by a process of translating natural language questions into database queries in languages such as SQL and SPARQL [22,23]; however, the coverage

of these systems is limited, and this has led to the exploration of neural network-based approaches [24] that promise more coverage. Perhaps the most widely-known application of question answering to knowledge graphs was the IBM Watson system [25], which used both "off-the-shelf", general-purpose knowledge graphs, as well as customized knowledge graphs to answer questions. Enterprise knowledge graphs are thus one of the key ways systems can be easily adapted to new domains, such as medicine [26]. As an example of the value that can be created in enterprises, the HAVASknowledge graph [27] can be considered a trendsetter. This knowledge graph describes start-up companies, and information about them was extracted from a variety of sources including existing structured data, data collected from social media and news events. This knowledge graph can also be updated and revised by local rapporteurs and other agents.

In fact, the process of knowledge graph refinement [28] is one of the key challenges for knowledge graphs and is still under-researched. There are primarily two kinds of refinement that can be applied to existing knowledge graphs: completion, which involves adding more facts and builds on techniques for extracting a knowledge graph [29], and error detection, which finds errors and (possibly) suggests corrections. A number of methods have been proposed for detecting errors in knowledge graphs: reasoning and the structure of the graph can be used to infer erroneous types or values [30]; incorrect links in the knowledge graph can be detected as outliers by statistical tests [31]; or external general-purpose knowledge can be exploited to find errors in enterprise knowledge graphs [32,33]. Moreover, an important part of knowledge graph refinement is the development of user interfaces that allow human users to easily visualize and verify the content of the knowledge graph, and thus, a collaborative knowledge graph development methodology has recently been embodied in tools such as VocBench [34].

3. Conversation Concepts

3.1. Problem Description

With increasing expectations for virtual assistants or chatbots, there is a need to empower chatbots with overall domain knowledge, contextual awareness within a conversation or chat session, followed by a personalized answer. A predefined answer template for a given question may not be adequate. A typical Q&A-based chatbot provides the same answer to every user for a question like "May I contribute to my retirement account?". However, in reality, the presence of the word "my" demands that the chatbot needs to respond differently based on the user profile. The answer may be different for an already retired user vs. a non-retired user.

To address this issue, we developed a tool named Conversational Concept AI (CCAI), based on the Saffron tool described in Section 4. This tool has been deployed internally within FMR LLC since September 2019. The goal of this tool is to enable our virtual assistant with a Natural Language Understanding (NLU) component that can perform:

- Intent identification, as well as entity detection for the dialogue manager
- Customer intent disambiguation by leveraging other metadata
- Dialogue pattern monitoring in real time to help provide personalized answers

These goals of the Conversation Concepts AI (CCAI) system were realized by the topic extraction and taxonomy generation from unstructured conversational text. Thus, we created a self-service text analytic tool trained on financial domain data that understands financial concepts like mutual fund or dividend reinvestment as an entity in a knowledge graph rather than a string, like most commercial products do, and that is able to provide a summary of the topics present in documents. The further tools can use this knowledge graph as a key part of the information in their workflows. The CCAI and other similar tools can leverage knowledge graphs to further discover challenges in the digital experience of users, as well as the complexity of domain-specific terminologies in comparison with common languages, e.g., security vs. stock. With the discovery of a taxonomy from unstructured data, we are able to understand the interlinked contents and pattern of digital navigation in the financial domain, for example account, money transfer and trading. This

helps the content writers develop concise and action-driven contents with a link to the next step to reduce ambiguity in content navigation.

The CCAI tool can parse a corpus of conversational text and extract entities and the relationship between entities, which enables the end users to track various business needs such as product mentions, content gaps, and trending topics in customer interactions and help optimize the customer service process.

3.2. Use Cases

As part of our collaboration, a survey of the use cases for knowledge graphs within FMR LLC was extracted, in order to inform how the Saffron system could be exploited to make products for the financial business:

Semi-supervised taxonomy builder: For many financial applications, there is a need to have a taxonomy of topics for simplified search and retrieval processes, organizing complex data with a hierarchical ordering based on business units, applications or user segments. As a result, multiple taxonomies across different products and applications evolved in silo developments. Over time, these overlapping taxonomies introduced inconsistencies, broken links and relationships and hence became difficult to manage. In particular, a need was discovered to merge multiple hand-crafted taxonomies and to automatically update these taxonomies as new contents or products are developed.

Multi-intent classifier: With the increased use of interactive digital assistants and advances in NLU technologies, users now expect chatbots to understand complex, multi-intent utterances. A single user utterance may contain context or a description of the issue followed by a request for help or information. Currently, our approach leverages a taxonomy of actionable terms, to identify multiple slots within a digital assistant conversation to detect multiple intents. Our approach also leverages co-reference resolution methods to disambiguate inter-linked entities between previous utterances.

The previous hypothetical example may be expressed as:

> User: Hi, I am thinking to transfer some money to that account but I am not sure if I am allowed to.

With the help of a domain knowledge graph/ontology, our objective is to identify entities and break down complex utterances into "context/description" and "request".

Real-time taxonomy mapping and state tracking: During the conversation through a virtual assistant, each utterance is analysed by an intent classifier, entity detector, co-reference resolution and pre-defined dialogue manager to understand the current state of the conversation to select the optimal response through the virtual assistant. For example, in a natural conversation, users often refer to previous utterances through pronouns. Therefore, a variation to the previous hypothetical example may look like:

> User: Hi, I am thinking to transfer some money to that account but I am not sure if I am allowed to.
>
> User: That's great, can you please provide instructions for wire transfer
>
> User: What is the routing number?

The above example may appear simple in human-human conversation; however, for a chatbot, the conversation is challenging due to cross-referencing ("that account"), inferred entities ("allowed to", "routing number") and the state transition of inter-linked entities (feasibility of wire transfer to a retirement account).

Our objective here is to map the entities in the utterances in real time, track the transition of intents through the traversal of a knowledge graph/ontology and provide the most relevant information to the user.

We take these use cases as being typical of the kind of use cases that knowledge graphs would have in the financial services sector. In addition, we identified some emerging use cases, including identifying topics or issues in open-ended customer feedback, email or search terms, and identified the change in natural trends due to seasonality and/or real-time events. These use cases may be of interest for future developments.

4. Topic and Taxonomy Extraction

Given the requirements for the CCAI tool, it was clear that current taxonomy extraction solutions were not suitable to realise its goals. As such, we further developed the Saffron system [2], to support the development of knowledge graphs within FMR LLC. In particular, key requirements that were identified were the extension of the automatic extraction system to cover more sophisticated relation types, as well as the development of tools to support the evolution of the knowledge graph. In this section, we describe the further developments that were made to meet the goals of the CCAI system.

4.1. Automatic Extraction

The CCAI tool builds on NLP components provided by the Saffron system [2]. This tool consists of a number of steps: firstly, term extraction, where terms are identified using parts-of-speech patterns and then ranked as relevant terms according to some metrics (for more details, see https://github.com/insight-centre/saffron/wiki/4.1.-Term-Extraction (accessed on 9/4/2021)). This typically requires that the term:

- contains a minimum and a maximum number of n-grams,
- does not start nor end with a stop word [35],
- follows a set of parts-of-speech patterns empirically found to be associated with terms [36],
- occurs within the corpus at least a given number of times.

Term extraction then uses a number of metrics to rank the candidate terms for their "termhood" by measures based on the frequency of the use of the terms, the context in which the terms occur and the relative frequency compared to some neutral background corpus. Secondly, there is a relation extraction system, which previously extracted hypernyms only using the methodology described in Sarkar et al. [37]; however, this has been extended to extract further relations (see below). Finally, there is a system that builds a taxonomy (a directed acyclic graph) from these relations. This is a hard problem, and for the moment, we approached this using a simple greedy approach, which we found produced a near-optimal answer in most cases.

4.2. Semi-Automatic Taxonomy Development

Automatic taxonomy development provides the benefit of decreasing the effort required to create taxonomies. However, the low link precision [19] reported by such algorithms creates a major barrier for their deployment in customer-facing enterprise systems. The wrong term or relation brings a risk associated with branding and the business itself. In order to address this problem, we developed a human-in-the-loop approach where experts in the domain of the enterprise can validate the resulting taxonomy and where the validation step can be used to further improve the automatic taxonomy development.

The approach used for taxonomy development with human-in-the-loop is shown in Figure 1. First, we used an approach for automatic taxonomy development based on the terms extracted. Next, this taxonomy was manually evaluated by one or more domain experts, where terms and relations between terms were assigned as valid or invalid. The changes performed during the validation step were immediately reflected on the taxonomy provided as output, but they could also be used to inform a new cycle of term extraction and taxonomy development.
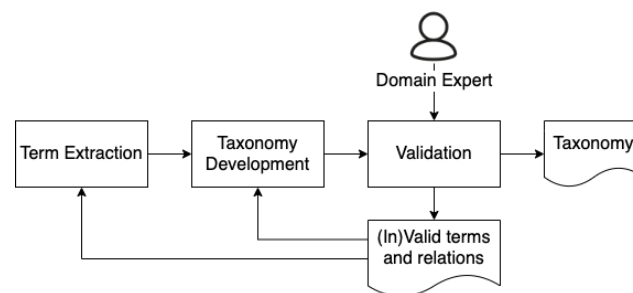
Figure 1. Human-in-the-loop taxonomy development pipeline.

In the taxonomy development step, we used the approach provided by Saffron [2]. However, any other algorithm could be potentially used under the requirement that it allows the enforcement of specific relations between terms.

For the validation step, we developed a user interface where a domain expert can validate each term and each taxonomic relation between terms.

Initially, each term is considered under revision, then a domain expert may change its status to accepted or rejected. When taxonomy retraining is triggered, each accepted term is enforced to be part of the new taxonomy, and each rejected term is removed before a taxonomy is constructed. Due to the removal of terms, new terms may now be considered as relevant enough by the term extraction algorithm to appear in the new taxonomy.

In addition to terms, domain experts are also allowed to either:

(i) approve a given parent-child relation or
(ii) change the parent for any given term within the taxonomy.

When a relation is approved or changed, the status of the terms involved in that relation are automatically changed to accepted. When the validation of relations is finished and the taxonomy retraining is triggered, all approved relations are included in the new taxonomy before any other possible relations are considered. Note also that, in order to enable the revision of the root term within the taxonomy, we added a virtual root node at the top of the taxonomy. Such a node is just a place holder at the top of each taxonomy, and it has no meaning as a term.

The resulting taxonomy, after one or multiple iterations of taxonomy development, can then be deployed within an application under the understanding that it has been validated by internal domain experts.

4.3. Extending Relation Type Extraction

As we wished to better understand the concepts extracted by the tool, we also developed a novel methodology for the extraction of generic relations from text. As with our previous works [37], we believed that there was sufficient information in the word embeddings in order to extract the relation between a pair of terms.

In order to train this model, we used ConceptNet [38], which is a knowledge graph containing general domain entities and relationship types between these entities. For the purpose of this work, we chose three relationship types from ConceptNet: "IsA", "PartOf", and "Synonym". "IsA" represents hypernymy relations; "PartOf" represents meronymy relations; and "Synonym" represents synonymy relations. In order to also generate a dataset for hyponymy relations, we reversed the direction of relations marked as hypernyms. All other relations appearing in ConceptNet were aggregated under the category "other". In total, there were 3.65 million relations, of which 11.1% where hypernym relations, 6.1% synonymy, 0.4% meronymy and the remaining 80.9% classified as "other". We extended our classifier using these data, so that Saffron is now one of the few open-source relation extraction systems that can build taxonomies with multiple relation types from text corpora.

4.4. Integration of Knowledge Graphs

Another key use case was the integration of knowledge graphs coming from different sources. This can include individual knowledge graphs in an organization, as well as those from external sources such as Wikidata (https://www.wikidata.org/ (accessed on 9/4/2021)), DBpedia [39] and YAGO [40]. We used the Naisc tool [41] (https://github.com/insight-centre/naisc (accessed on 9/4/2021)) that links multiple resources for this task. The structure of this tool is demonstrated in Figure 2 above, where two taxonomies were taken as the input, and then, a sequence of tools operated on the taxonomies to provide an accurate linking:

1. Analysis of the taxonomies to find the structure of the model;
2. Blocking to find (with reduced computational cost) a set of candidate links;
3. Pre-linking of those elements where this can easily be achieved, for example because there is only one candidate to link to;
4. Lenses to extract relevant textual elements;
5. Text features applying state-of-the-art NLP techniques to extract numerical measures of similarity;
6. Graph features analysing the graph and finding similar links;
7. A scorer analysing each link and finding the likelihood of a link being established;
8. Rescaling used to fix the link quality;
9. A matcher using "common-sense" constraints to find the most likely global matching.
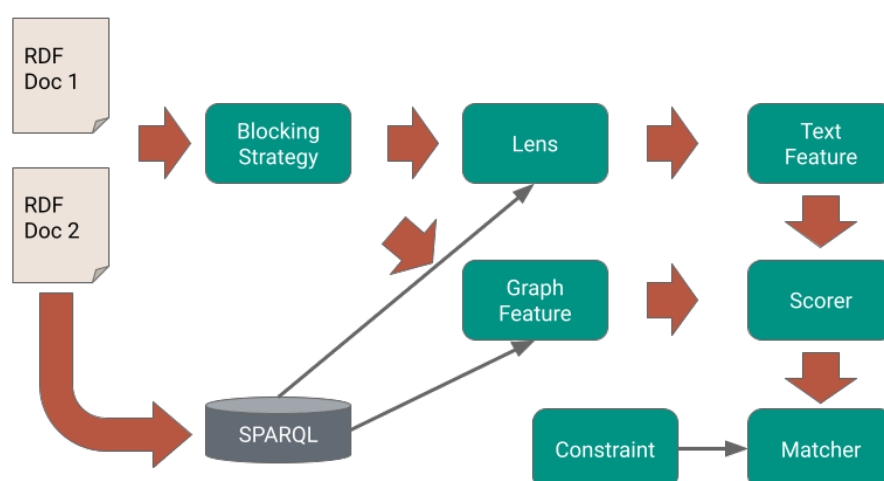

Figure 2. The architecture of the Naisc system for taxonomy linking.

As a final output, the system outputs a set of links between the two taxonomies, stating the equivalence between; these can then be used to build a single knowledge graph from multiple sources.

4.5. Deployment and Future Plans

Using the techniques developed in the previous section, the CCAI tool was deployed, and examples of its application are shown in Figures 3 and 4. In Figure 3, we see the terms that were extracted by the tool in the order of frequency, showing the user the main topics that were discussed within the corpus. Then, in Figure 4, we see an example of a taxonomy that was constructed, where users can expand individual nodes by clicking on the bubbles in order to navigate the knowledge graph.
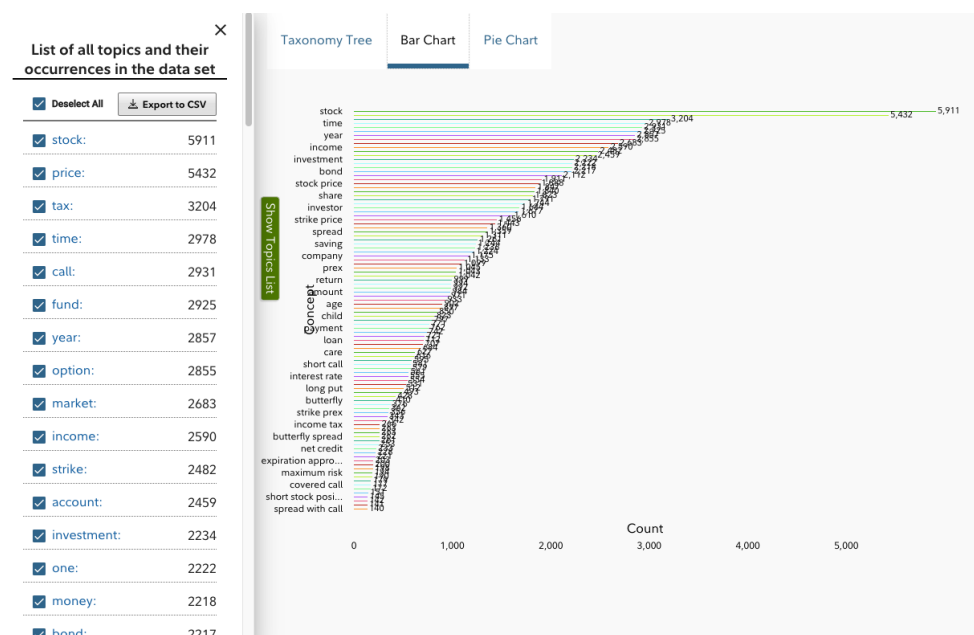
Figure 3. A screenshot of the Conversation Concepts Artificial Intelligence (CCAI) tool, showing the topics extracted from the financial articles (LC). LC documents cover a range of financial education topics and are publicly available. The key terms are ordered by the frequency of their occurrence within the corpus.
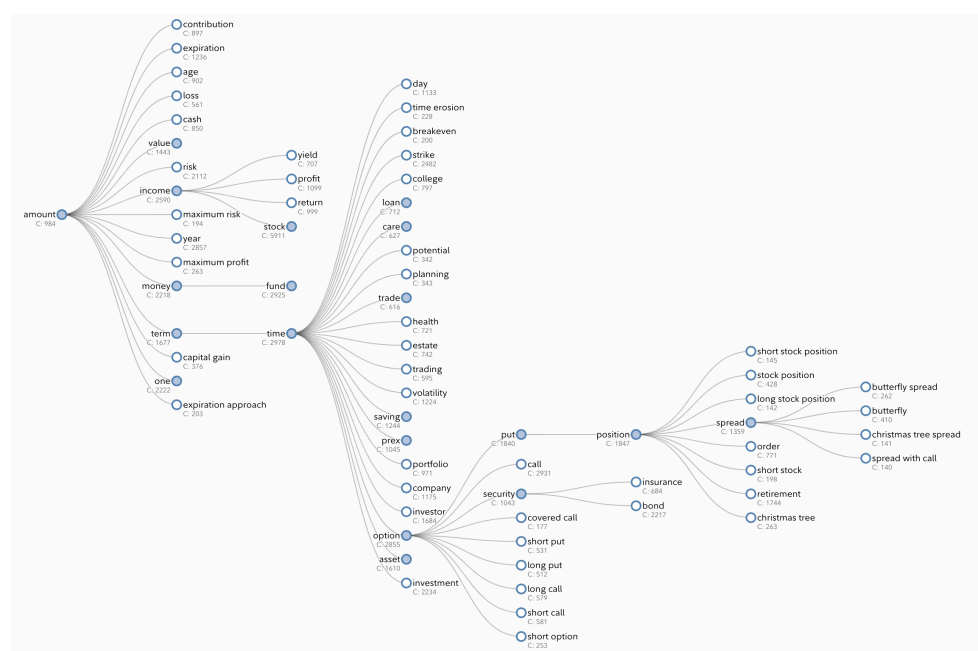

Figure 4. A partial taxonomy generated by the CCAI tool and Saffron, extracted from LC documents. The key terms in this taxonomy are the same as those shown in Figure 3. Some nodes are collapsed (filled circles) to facilitate visualization as a figure.

The CCAI tool outputs one or more taxonomies based on the source of the unstructured data. These outputs then help us to discover the pattern of digital content with respect to particular financial product/services. This pattern analysis is a critical component in designing a dialogue manager, entity linking or query disambiguation, as well as optimizing our digital contents. Without the tool, the mentioned task may require substantial manual effort to scale our virtual assistant and content discovery tools.

Another natural extension of CCAI's capability is to enable a personalization concept by overlaying an observed pattern with context. With the current development of a context-based language model like BERT [42], CCAI can be extended to understand the context within a conversation as well.

5. Discussion

The enterprise of the future will be self-reflective, making it smarter by exploiting AI in gaining a broader and deeper understanding of its operations, its customers, employees and operating context, driving better services and better performance. The cognitive industrial revolution that is now unfolding builds on technology that understands the world in which it operates in ways that are not dissimilar to human understanding. AI enables the development of technology with cognitive abilities, from listening, reading and seeing to deciding and reacting. We already see applications of this on our smartphones with digital assistants, machine translation and face recognition, but this is only the very start. Making machines smarter is only part of this (see also https://www.irishtimes.com/business/work/this-is-the-revolutionary-age-of-machines-that-can-understand-1.3507349 (accessed on 9/4/2021)).

The biggest impact will be when human collectives, such as the enterprise, develop cognitive abilities that enable fully automatic, real-time and context-sensitive analytics on all possible types of enterprise data, from numerical, textual and visual to other emerging types of data, including sensory. The resulting analytic profile of the enterprise will be more dynamic and more complete, in turn enabling informed decision making by the enterprise itself, as if it were a cognitive entity. The cognitive enterprise, as advocated by IBM, aims to develop a new generation of enterprise information system that integrates AI in order to "understand unstructured data, reason to form hypotheses, learn from experience and interact with humans naturally" (https://www.ibm.com/watson/advantage-reports/getting-started-cognitive-technology.html(accessed on 9/4/2021)). This vision builds primarily on the initial success of Watson, the intelligent dialogue system that famously won Jeopardy and that has since been developed further to be deployed in health care and other domains [43].

A central aspect of this vision is access to an underlying knowledge graph, which enables the reasoning process, as well as provides the context for data analysis, machine learning and system-to-human interaction. As discussed in the Introduction, the objective of enterprise knowledge graph development has therefore now been adopted by many companies, from IBM and Google to Facebook, Microsoft and eBay [44], among many others [2]. In this paper, we outlined an approach towards the automatic development of knowledge graphs from enterprise data. Our approach is generic and can be applied to any enterprise setting with large amounts of unstructured textual data.

6. Conclusions

In this work, we presented how the open-source taxonomy extraction system Saffron and taxonomy linking system Naisc, developed by NUI Galway, were applied within the Conversational Concepts Artificial Intelligence application for the financial domain. We applied these to the use cases and demonstrated that high-quality text extraction applications can be converted into useful applications due to the wide applications of the enterprise knowledge graphs. This was elucidated by the application in not only taxonomy construction, but also secondary applications such as intent classification and dialogue analysis, with further applications including customer satisfaction analysis. As such, we see this application as a key enabler of the idea of a "smart enterprise" and expect that similar results can be obtained by other companies who adopt the methodologies employed in this paper. We expect that these use cases and our solution to these use cases will be of value to others who wish to solve similar problems in financial services and other related domains.

References

1. Gomez-Perez, J.M.; Pan, J.Z.; Vetere, G.; Wu, H. Enterprise knowledge graph: An introduction. In *Exploiting Linked Data and Knowledge Graphs in Large Organisations*; Springer: New York, NY, USA, 2017; pp. 1–14.
2. Pereira, B.; Robin, C.; Daudert, T.; McCrae, J.P.; Buitelaar, P.; Mohanty, P. Taxonomy Extraction for Customer Service Knowledge Base Construction. In *Proceedings of the SEMANTicS 2019*, Springer: New York, NY, USA,; 2019.
3. Quillian, M.R. Word concepts: A theory and simulation of some basic semantic capabilities. *Behav. Sci.* **1967**, *12*, 410–430.
4. Shadbolt, N.; Berners-Lee, T.; Hall, W. The Semantic Web revisited. *IEEE Intell. Syst.* **2006**, *21*, 96–101.
5. Bizer, C.; Heath, T.; Berners-Lee, T. Linked data: The story so far. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts*; IGI Global: Hershey, PA, USA, 2011; pp. 205–227.
6. Brickley, D.; Guha, R.V.; Layman, A. *Resource Description Framework (RDF) Schema Specification*, https://www.w3.org/TR/rdf-schema/ (accessed on 9/4/2021); 1999.
7. McGuinness, D.L.; Van Harmelen, F.; others. OWL web ontology language overview. *W3C Recomm.* **2004**, *10*, 2004.
8. Singhal, A. Introducing the Knowledge Graph: Things, not Strings. 2012. Available online: https://www.blog.google/products/search/introducing-knowledge-graph-things-not/ (accessed on 9/4/2021).
9. Penela, V.; Álvaro, G.; Ruiz, C.; Córdoba, C.; Carbone, F.; Castagnone, M.; Gómez-Pérez, J.M.; Contreras, J. miKrow: Semantic Intra-enterprise Micro-Knowledge Management System. In *The Semantic Web: Research and Applications*; Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 154–168.
10. Denaux, R.; Ren, Y.; Villazon-Terrazas, B.; Alexopoulos, P.; Faraotti, A.; Wu, H. Knowledge Architecture for Organisations. In *Exploiting Linked Data and Knowledge Graphs in Large Organisations*; Springer: New York, NY, USA, 2017; pp. 57–84.
11. Ngomo, A.C.N.; Auer, S.; Lehmann, J.; Zaveri, A. Introduction to linked data and its lifecycle on the web. In *Reasoning Web International Summer School*; Springer: New York, NY, USA, 2014, pp. 1–99.
12. Buitelaar, P.; Cimiano, P.; Magnini, B. *Ontology Learning from Text: Methods, Evaluation and Applications*; IOS Press: Amsterdam, The Netherlands, 2005 ; Volume 123.
13. Astrakhantsev, N. ATR4S: Toolkit with state-of-the-art automatic terms recognition methods in Scala. *Lang. Resour. Eval.* **2018**, *52*, 853–872.
14. Bordea, G.; Buitelaar, P. DERIUNLP: A context based approach to automatic keyphrase extraction. In *Proceedings of the 5th International Workshop on Semantic Evaluation*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 146–149.
15. Hearst, M.A. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics-Volume 2*; Association for Computational Linguistics: Stroudsburg, PA, USA, 1992; pp. 539–545.
16. Shwartz, V.; Santus, E.; Schlechtweg, D. Hypernyms under Siege: Linguistically-motivated Artillery for Hypernymy Detection; In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 65–75.
17. Roller, S.; Kiela, D.; Nickel, M. *Hearst Patterns Revisited: Automatic Hypernym Detection from Large Text Corpora*; In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018, Volume 2: Short Papers*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 358–363.
18. Wang, C.; Fan, Y.; He, X.; Zhou, A. A family of fuzzy orthogonal projection models for monolingual and cross-lingual hypernymy prediction. In *The World Wide Web Conference*, San Francisco, CA, USA, 13–17 May 2019; pp. 1965–1976.
19. Bordea, G.; Lefever, E.; Buitelaar, P. SemEval-2016 task 13: Taxonomy extraction evaluation (TExEval-2). In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 1081–1091.
20. Nayak, G.; Dutta, S.; Ajwani, D.; Nicholson, P.; Sala, A. Automated assessment of knowledge hierarchy evolution: Comparing directed acyclic graphs. *Inf. Retr. J.* **2019**, *22*, 256–284.
21. Diefenbach, D.; Lopez, V.; Singh, K.; Maret, P. Core techniques of question answering systems over knowledge bases: A survey. *Knowl. Inf. Syst.* **2018**, *55*, 529–569.
22. Unger, C.; Cimiano, P. Pythia: Compositional meaning construction for ontology-based question answering on the semantic web. In *International Conference on Application of Natural Language to Information Systems*; Springer: New York, NY, USA, 2011; pp. 153–160.

23. Lopez, V.; Fernández, M.; Motta, E.; Stieler, N. PowerAqua: Supporting users in querying and exploring the semantic web. *Semant. Web* **2012**, *3*, 249–265.

24. Lukovnikov, D.; Fischer, A.; Lehmann, J.; Auer, S. Neural network-based question answering over knowledge graphs on word and character level. In Proceedings of the 26th international conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1211–1220.

25. Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A.A.; Lally, A.; Murdock, J.W.; Nyberg, E.; Prager, J.; et al. Building Watson: An overview of the DeepQA project. *AI Mag.* **2010**, *31*, 59–79.

26. Ahmed, M.N.; Toor, A.S.; O'Neil, K.; Friedland, D. Cognitive computing and the future of health care cognitive computing and the future of healthcare: The cognitive power of IBM Watson has the potential to transform global personalized medicine. *IEEE Pulse* **2017**, *8*, 4–9.

27. Gutiérrez-Cuellar, J.; Gómez-Pérez, J.M. HAVAS 18 Labs: A Knowledge Graph for Innovation in the Media Industry. In *International Semantic Web Conference (Industry Track)*; Riva del Garda, Italy, 19–23 October 2014; Volume 1383.

28. Paulheim, H. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semant. Web* **2017**, *8*, 489–508.

29. Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning entity and relation embeddings for knowledge graph completion. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.

30. Nakashole, N.; Theobald, M.; Weikum, G. Scalable knowledge harvesting with high precision and high recall. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, 9–12 February 2011; pp. 227–236.

31. Paulheim, H. Identifying wrong links between datasets by multi-dimensional outlier detection. In Proceedings of the Third International Workshop on Debugging Ontologies and Ontology Mappings (WoDOOM), Anissaras/Hersonissou, Greece, 26 May 2014; pp. 27–38.

32. Lehmann, J.; Gerber, D.; Morsey, M.; Ngomo, A.C.N. Defacto-deep fact validation. In *International Semantic Web Conference*; Springer: New York, NY, USA, 2012; pp. 312–327.

33. Liu, S.; d'Aquin, M.; Motta, E. Towards Linked Data Fact Validation through Measuring Consensus. In *Proceedings of the 2nd Workshop on Linked Data Quality co-located with 12th Extended Semantic Web Conference (ESWC 2015), Portorož, Slovenia, 1 June 2015*; CEUR Workshop Proceedings No. 137, (2015).

34. Stellato, A.; Rajbhandari, S.; Turbati, A.; Fiorelli, M.; Caracciolo, C.; Lorenzetti, T.; Keizer, J.; Pazienza, M.T. VocBench: A web application for collaborative development of multilingual thesauri. In *European Semantic Web Conference*; Springer: New York, NY, USA, 2015; pp. 38–53.

35. Hulth, A. Enhancing linguistically oriented automatic keyword extraction. In *Proceedings of the HLT-NAACL 2004: Short Papers*; Boston, MA, USA, 2–7 May 2004; pp. 17–20.

36. Bordea, G. Domain Adaptive Extraction of Topical Hierarchies for Expertise Mining. Ph.D. Thesis, NUI Galway, Galway, Ireland, 2013.

37. Sarkar, R.; McCrae, J.P.; Buitelaar, P. A supervised approach to taxonomy extraction using word embeddings. In Proceedings of the 11th Language Resource and Evaluation Conference (LREC), Miyazaki, Japan, 7–12 May 2018.

38. Speer, R.; Chin, J.; Havasi, C. ConceptNet 5.5: An open multilingual graph of general knowledge. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.

39. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z.G. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web, Proceedings of the 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, 11–15 November 2007*; Lecture Notes in Computer Science; Aberer, K., Choi, K., Noy, N.F., Allemang, D., Lee, K., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Springer: New York, NY, USA, 2007; Volume 4825, pp. 722–735, doi:10.1007/978-3-540-76298-0_52.

40. Hoffart, J.; Suchanek, F.M.; Berberich, K.; Weikum, G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell.* **2013**, *194*, 28–61, doi:10.1016/j.artint.2012.06.001.

41. McCrae, J.P.; Buitelaar, P. Linking Datasets Using Semantic Textual Similarity. *Cybern. Inf. Technol.* **2018**, *18*, 109–123, doi:10.2478/cait-2018-0010.

42. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arxiv:1810.04805.

43. Ferrucci, D.; Levas, A.; Bagchi, S.; Gondek, D.; Mueller, E.T. Watson: Beyond Jeopardy! *Artif. Intell.* **2013**, *199*, 93–105.

44. Noy, N.; Gao, Y.; Jain, A.; Narayanan, A.; Patterson, A.; Taylor, J. Industry-scale knowledge graphs: Lessons and challenges. *Queue* **2019**, *17*, 48–75.