IVEA: An Information Visualization Tool for Personalized Exploratory Document Collection Analysis

VinhTuan Thai, Siegfried Handschuh, and Stefan Decker

Digital Enterprise Research Institute (DERI),
National University of Ireland, Galway
`firstname.lastname@deri.org`

Abstract. Knowledge work in many fields requires examining several aspects of a collection of documents to attain meaningful understanding that is not explicitly available. Despite recent advances in document corpus visualization research, there is still a lack of principled approaches which enable the users to personalize the exploratory analysis process. In this paper, we present IVEA (**I**nformation **V**isualization for **E**xploratory Document Collection **A**nalysis), an innovative visualization tool which employs the PIMO (Personal Information Model) ontology to provide the knowledge workers with an interactive interface allowing them to browse for information in a personalized manner. Not only does the tool allow the users to integrate their interests into the exploration and analysis of a document collection, it also enables them to incrementally enrich their PIMO ontologies with new entities matching their evolving interests in the process, and thus benefiting the users not only in their future experiences with IVEA but also with other PIMO-based applications. The usability of the tool was preliminarily evaluated and the results were sufficiently encouraging to make it worthwhile to conduct a larger-scale usability study.

Keywords: Personal Information Management, Semantic Desktop, PIMO ontology, Exploratory Document Collection Analysis, Information Visualization, Coordinated Multiple Views, Human-Computer Interaction.

1 Introduction

Apart from the need to retrieve information from documents that are relevant to certain topics of interest, oftentimes knowledge workers also need to explore and analyze a collection of documents as a whole to gain further understanding. This particular kind of information-seeking activity can be referred to as *exploratory data analysis* or *information analysis* and is commonly carried out in science, intelligence and defense, or business fields [1]. Unlike the information retrieval activity, the information analysis activity aims to provide the users with an overall picture of a text collection as a whole on various dimensions instead of presenting them with the most relevant documents satisfying some

search criteria. The insights obtained from the exploration and analysis of a text collection can enable the knowledge workers to understand the distribution of topics, to find clusters of similar documents, or to identify trends or linkages between different entities [1]. Information visualization is an effective mechanism to support the information analysis task and has also been widely employed in many data mining and knowledge discovery tools to identify previously unknown useful information.

While there are many existing tools reported in the Information Visualization literature that support document corpus visualization (such as [1,2,3,4]), most are based only on the text of the documents in the corpus and hence present findings that are independent of the users' interests. Their approaches focus on the entity extraction process in order to identify the main entities (e.g. topics, people, locations) within a document collection and then visualize different linkages between the identified entities onto a 2D or 3D user interface. While the automatic extraction of entities is helpful, the visual exploration process cannot be aligned with the knowledge workers' interests, especially when a number of entities contained within their spheres of interests are of particular importance to their knowledge work. As a result, the knowledge workers cannot have a personal viewpoint over the entities and relationships that they wish to focus on to gain insights by using the existing tools. There is a clear need to link various important concepts and structures within parts of the knowledge workers' mental models with the document collection visual exploration and analysis activity. To achieve this, it is necessary that these concepts and structures can be externalized to some formal representation.

The above requirement motivates us to investigate existing work within the Semantic Desktop area where several integrated Personal Information Management environments have been introduced [5]. In the Semantic Desktop paradigm, desktop items are treated as Semantic Web resources and formal ontologies "allow the user to express personal mental models and form the semantic glue interconnecting information and systems" [5]. One such ontology has been designed and developed within the Gnowsis project and is based upon the *Personal Information Model* (PIMO) [6]. Since the PIMO ontology acts as a "formal representation of the structures and concepts" within a knowledge worker's mental model [6], it can be employed as a means to integrate her personal interests into an exploratory visualization tool. This would give the PIMO-based visualization tool clear advantages over the keyword-based visualization tools in that: (1) it can align the exploration process with the users' interests and hence can offer the users the control over which aspects of a text collection they wish to focus on, (2) it can utilize the hierarchical structure (class-subclasses, class-instances) within the PIMO ontology to provide the users with the flexibility to explore a text collection at different levels of detail.

In this context, we propose an innovative visualization tool called IVEA, which leverages upon the PIMO ontology and the Coordinated Multiple Views technique to support the personalized exploration and analysis of document collections. IVEA allows for an interactive and user-controlled exploration process in

which the knowledge workers can gain meaningful, rapid understanding about a text collection via intuitive visual displays. Not only does the tool allow the users to integrate their interests into the visual exploration and analysis activity, it also enables them to incrementally enrich their PIMO ontologies with entities matching their evolving interests in the process. With the newly added entities, the PIMO ontology becomes a richer and better representation the users' interests and hence can lead to better and more personalized exploration and analysis experiences in the future. Furthermore, not only can IVEA be beneficial to its targeted task, it provides an easy and incremental way that requires minimal effort from the users to keep their PIMO ontologies in line with their continuously changing interests. This, indirectly, can also benefit other PIMO-based applications.

The remainder of this paper is organized as follows: Section 2 details different visual aspects of IVEA. Its system architecture is described in Section 3. In Section 4, we present a usability study and a discussion on the results. Related work is highlighted in Section 5. Section 6 concludes the paper and outlines future work.

2   Proposed Solution

2.1   Design Desiderata

In order to design a suitable visualization solution, various factors need to be taken into consideration, including: (1) *the nature of the task* (personalized exploratory analysis, integration of newly discovered entities to evolving spheres of interests), (2) *the type of data* (contents of documents in a collection rather than those of a single metadata file describing it), (3) *the target users* (knowledge workers rather than computer users in general), and (4) *their typically available work environment* (2D display interface rather than advanced 3D environment). Furthermore, the design of advanced visual interfaces can also benefit from the well-known visual information-seeking mantra: "*Overview first, zoom and filter, then details-on-demand*" proposed by Shneiderman in his seminal paper [7]. Taking into account all those aspects, desirable capabilities for such a visualization solution would be:

- Provide an overall view showing the degrees of relevance of all documents in a collection with respect to the personal interests represented by the PIMO ontology.
- Enable the users to interactively explore and analyze a text collection at different levels of detail by filtering the set of documents based on the PIMO concepts and instances and their hierarchical relationships.
- Provide a detailed view on demand which focuses on the distribution of entities of interest within each document.
- Suggest the users with entities potentially matching their evolving interests and enable them to enrich the PIMO ontology with the selected entities.

2.2   Visual Interface

To cater for the above-mentioned needs, we employ the Coordinated Multiple Views (CMV) technique [8] to design IVEA, as shown in Fig. 1. The use of the CMV technique enables the users to understand a text collection better via different visual displays highlighting different aspects of it, as well as via the interactions with and coordinations between those displays. Interested readers are encouraged to view the demo screencast available online[1].
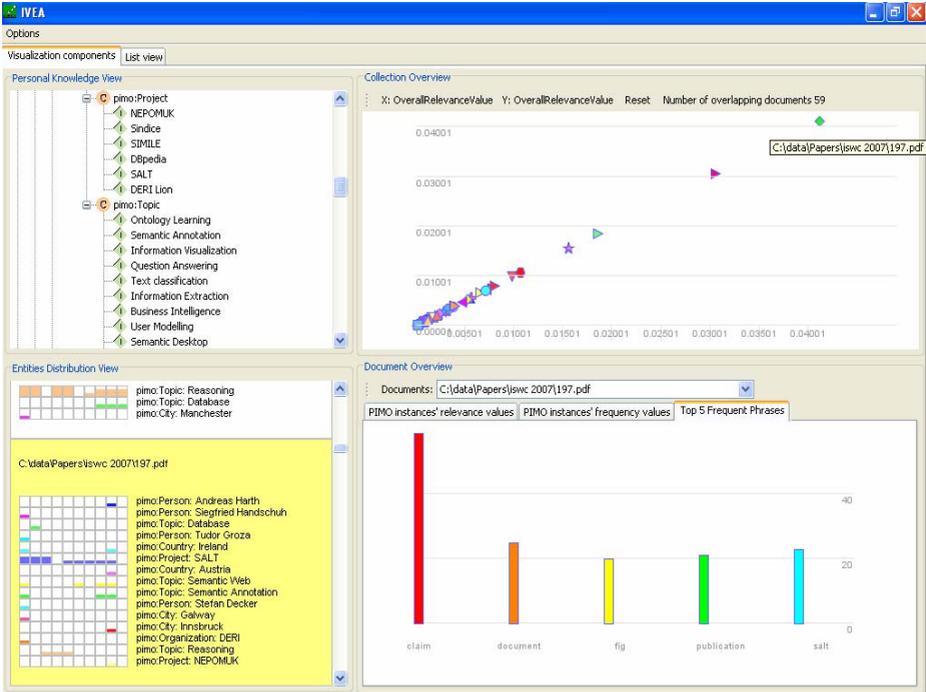

Fig. 1. IVEA interface

IVEA's visual interface consists of four views as follows:

– **Personal Knowledge View**: The tree structure, as shown on the upper-left corner of Fig. 1, is used to display the concepts, instances and their hierarchical relationships within the PIMO ontology. Although this tree structure does not display many other relationships that exist between PIMO concepts and instances, it is sufficient to serve as an anchor keeping the users informed of their entities of interest. This view acts as the basis for some of the interactions in IVEA.

---

[1] http://smile.deri.ie/projects/ivea/demo/demo1.html

- **Collection Overview:** The scatter plot, as shown on the upper-right corner of Fig. 1, is used as the overall view to display documents matching the users' interests. On the scatter plot, each document is represented by a dot and its file name is shown in the dot's tooltip text. The coordinate of a dot on the scatter plot is determined by the relevance values of the dot's corresponding document with respect to the classes or instances set on the x and y axes. The initial display of the scatter plot, as shown in Fig. 1, uses the overall relevance values of documents in the collection with respect to the set of all PIMO instances on both axes. Based on this initial display, the users can, for example, see that in the collection being explored, 59 documents overlap with their interests and that the document "$C:\backslash data\backslash Papers\backslash iswc$ $2007\backslash 197.pdf$", represented by the rightmost dot on the scatter plot, is most relevant, based on its coordinate. More details about that particular document can be obtained immediately from the coordinated document overview and entities distribution view as described shortly. Moreover, the dimension of either of the two axes can be changed to reflect how relevant the documents are with respect to the concepts or instances placed on it. The dots' colors and shapes are used to differentiate the associated documents. The dot's size can be customized to accommodate for text collections of different sizes.

- **Document Overview:** Bar charts are used to display detailed views on various characteristics of each document in the collection. Three different bar charts are used on the lower-right corner of Fig. 1. The first bar chart shows the PIMO instances appearing in a document together with the relevance values of that document with respect to them. The second one displays the matching PIMO instances based on their frequencies. The third bar chart shows the top frequent phrases or terms in a document. For instance, in Fig. 1, the top frequent phrases bar chart of the document "*197.pdf*" mentioned above shows that the word "*claim*" appears most frequently.

- **Entities Distribution View:** This view, as shown in the lower-left corner of Fig. 1, is based on the TileBars, originally reported in [9], and some of its variants in [10]. It is used to display the matching PIMO instances within each fragment of a document. The rows are associated with PIMO instances whose labels are placed next to them on the right. The number of column in this view is the number of fragments to divide a document up to, whose value can be set by the users. Each document is split into sentences and they are put into fragments such that each fragment contains an approximately equal number of sentences. The height of the colored area in a cell is determined by the frequency of the corresponding PIMO instance in that particular fragment. By using this view, the users can quickly be informed of the relative locations, in which certain entities of interest appear, together with their respective frequencies. For instance, Fig. 1 shows that the instance "*SALT*" of the concept "*pimo:Project*" appears more often in the first three fragments than in any other parts of the document "*197.pdf*".

2.3   Interactions, Manipulations and Coordinations

In line with the visual information-seeking mantra [7], IVEA provides the users with the freedom to interact and control the visual displays in the following manners:

– **Filtering:** It is essential that the users are able to filter out uninteresting documents to focus only on a restricted subset of a text collection. In IVEA, the users can directly manipulate the overall view displayed in the scatter plot by dragging concepts or instances from the personal knowledge view and dropping them onto the labels of the x and y axes to indicate the dimensions upon which the relevance values of documents are to be measured. IVEA instantly updates the scatter plot by executing dynamic queries for the relevance values of all documents with respect to the instance or the aggregated relevance values with respect to the concept placed on the axes. In Fig. 2, the x-axis highlights documents relevant to "*Semantic Desktop*" while the y-axis highlights documents relevant to "*pimo:Person*". The relevance value of a document with respect to "*pimo:Person*" is dynamically measured as the aggregated relevance value of that document with respect to all instances of the concept "*pimo:Person*" in the PIMO ontology. Hence, the scatter plot can show, among others, documents referring to both the topic "*Semantic Desktop*" and one or more persons who are of specific interest to the users (documents plotted above both axes). The example also illustrates that IVEA can take advantage of the hierarchical relationships between entities within the PIMO ontology to allow for the rapid exploration of a text collection at different levels of detail.
– **Details-on-demand:** Once the overall view has been restricted to a specific set of documents, the users can investigate the details of each document within that set. Clicking on a dot in the scatter plot will open up the corresponding document in its associated application. Furthermore, to interactively link the collection overview with the detailed views, coordinations between the scatter plot and the bar charts as well as the Entities Distribution View are provided in such a way that hovering the mouse over a dot in the scatter plot will display the corresponding bar charts and highlight the TileBars of the document represented by that dot. This interaction is also illustrated in Fig. 2.
– **PIMO ontology enrichment:** An innovative feature of IVEA is its capability to enable the users, with minimal efforts, to enrich their PIMO ontologies while exploring a text collection. While investigating a particular document to see how it overlaps with their spheres of interests, the users are presented with the most frequent phrases[2] in that document. The top frequent phrases and their corresponding frequencies are displayed in a histogram as shown in Fig. 3. If any of the presented phrases is of specific interest to the users, they are just two-click away from adding it to the PIMO ontology simply by dragging its respective column in the histogram and dropping on a concept

---

[2] Candidate phrases are nouns or noun chunks consisting of at most 3 words.
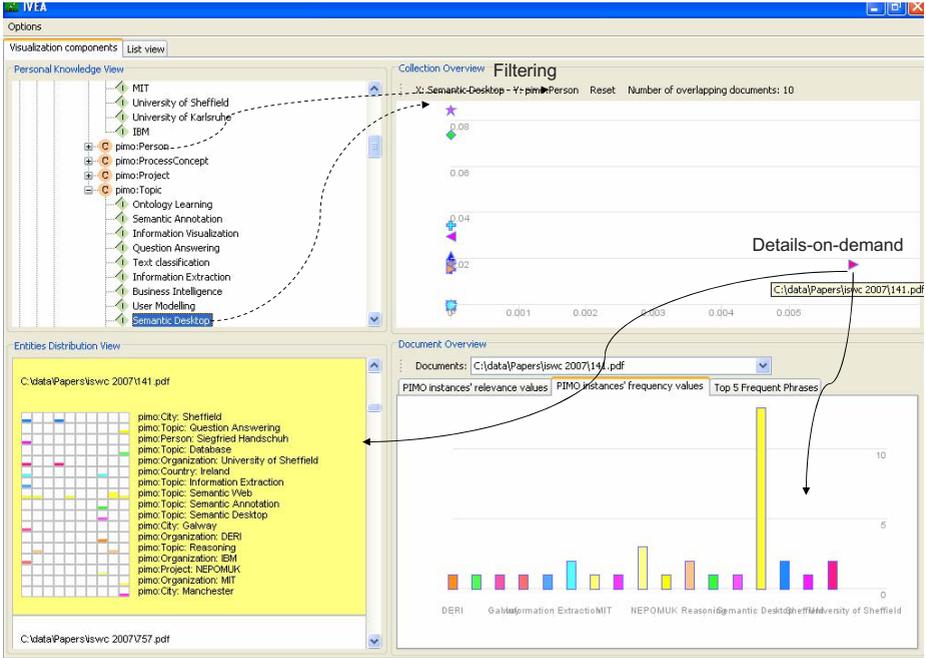
Fig. 2. Interactive Filtering and Details-on-demand

in the PIMO ontology. The users have the option to add the selected phrase as a subclass or instance of the targeted concept. Fig. 3 and 4 illustrate how the users can enrich the PIMO ontology by adding "*conference*" as a subclass of "*pimo:Event*". We believe that this is of benefit to the users as it allows them to update their PIMO ontologies with new entities on-the-fly. Consequently, they can better explore a text collection when their spheres of interests are better represented. Besides, an extended PIMO ontology is useful not only for IVEA but also for other PIMO-based applications.

3   System Architecture

The system architecture is shown in Fig. 5. Documents in a collection are first analyzed by the text processing component, which is based on various natural language processing resources (Tokenizer, Sentence Splitter, POS Tagger, JAPE Transducer) provided by the GATE [11] API. Each document is split into sentences to identify the fragments' boundaries for use in the Entities Distribution View. Furthermore, to suggest the users with entities potentially matching their interest, most frequent noun chunks in each document together with their frequencies are extracted.
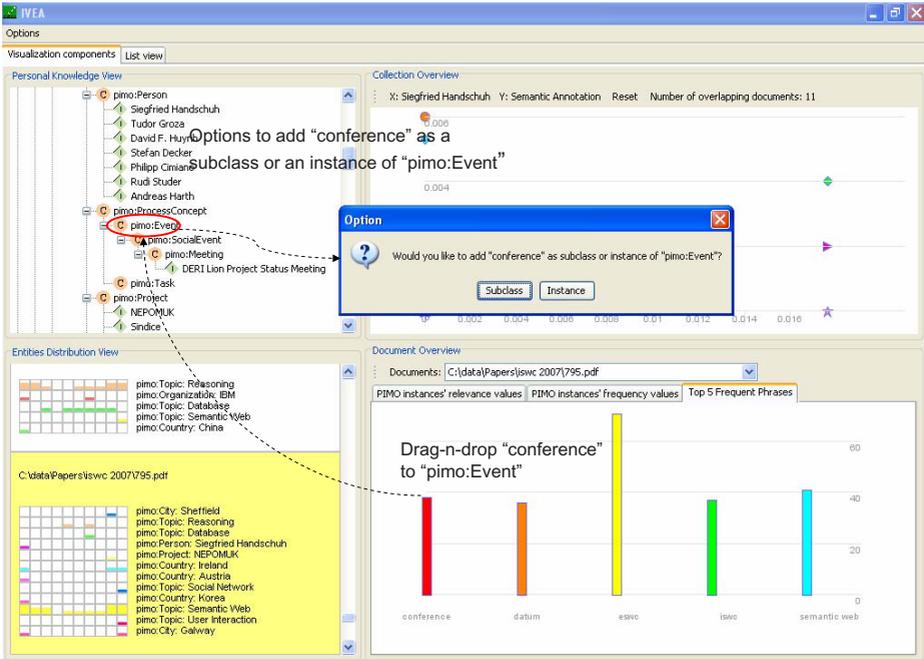
Fig. 3. Adding new concepts or instances to the PIMO ontology
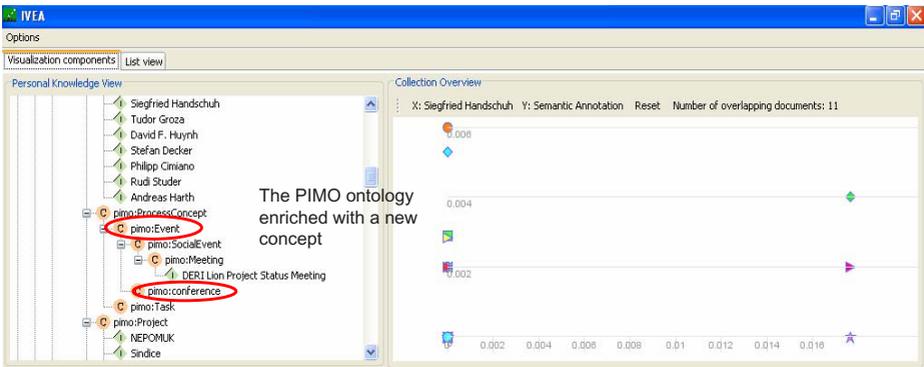
Fig. 4. The PIMO ontology added with a new concept

The analyzed text is then stored into a Lucene[3] index. A Boolean query (with the default OR operator) consisting of all PIMO instances is used to retrieve documents that are of interest to the users. The term weight in a document of an instance is used as the relative relevance value of that document with respect

[3] http://lucene.apache.org/
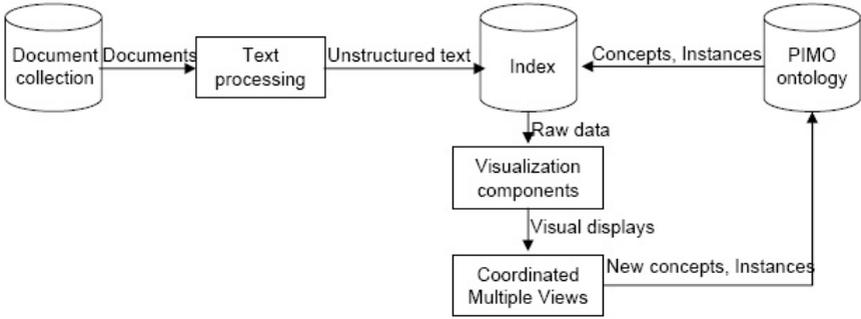
Fig. 5. System architecture

to that instance. In Lucene, a variant of the well-known TF.IDF term weight function is used, which takes into account the frequency of a term locally (in a document) and globally (in the whole collection), as well as the length of a document itself [12]. The relative relevance value of a document with respect to a class is the aggregated relevance value of that document with respect to all of its direct instances and recursively, all of its subclasses.

The above relevance values and the frequencies of PIMO instances in fragments of each document are used as raw data for the visualization components. The implementation of the visual displays is based on the prefuse library [13] and Java Swing components.

4   Usability study

Given a text collection on the desktop, most users are used to keyword-based search interface to look for documents satisfying some criteria using a desktop search engine. To explore a text collection this way, the users may have to perform many searches, each with a query containing multiple Boolean clauses. As such, we implemented a baseline interface as shown in Fig. 6 and performed a small-scale usability study to (1) get initial indications about the users' performances by using IVEA versus by using the baseline interface, and (2) identify IVEA's potential usability problems.

4.1   Description

Six researchers participated in the study. They all had prior knowledge about ontology and Boolean query operators. They were assumed to have the same sphere of interest, whose concepts and structures were encoded into a predefined PIMO ontology. This PIMO ontology acted as the basis for exploring and analyzing a test collection consisting of 62 research papers in the Semantic Web domain. Although the collection's size was not large, it was suitable for this initial study since it could
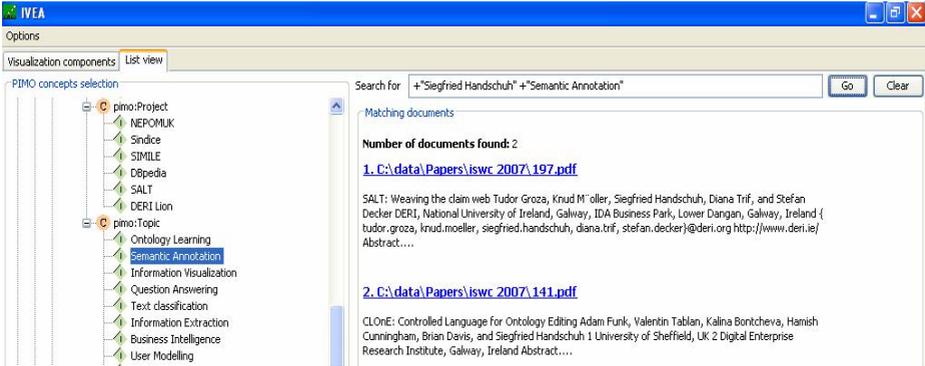
Fig. 6. Baseline interface

be representative of some real-world situation, e.g. when a researcher wishes to understand different characteristics of a document collection consisting of scientific papers published at a conference in a particular year.

The subjects were asked to perform the same set of 4 simple tasks and 4 complex tasks by using first the baseline interface and then the visual interface, followed by filling out the follow-up questionnaire. The following measures were recorded based on the subjects' responses:

- **Effectiveness**: calculated as the accuracy and completeness with which the users accomplished a task's goal [14].
- **Temporal efficiency**: calculated as the ratio between the effectiveness achieved in a task and the time taken to complete that task [14].
- **Users' satisfaction**: determined by the ratings responded to the follow-up questionnaire consisting of 12 Likert-scale questions (the ratings are on a range from -2 (strongly disagree/very bad) to 2 (strongly agree / very good)) and 1 open-ended question seeking suggestions on improving IVEA's design.

Further details about the tasks, the questionnaire and their respective results can be found online[4]. In the next section, we elaborate our findings and discuss what can be deduced from them.

4.2   Findings

We are aware that the study described above is not a summative usability evaluation, which would (1) involve a larger number of participants who are representative of the target users population (knowledge workers) and hence may have limited knowledge about ontology and Boolean query operators, (2) require a larger test collection, (3) use a more extensive set of tasks, and (4) require that each interface be used first by half of the subjects. Nevertheless, at this stage

---

[4] http://smile.deri.ie/projects/ivea/usability/preliminary.html
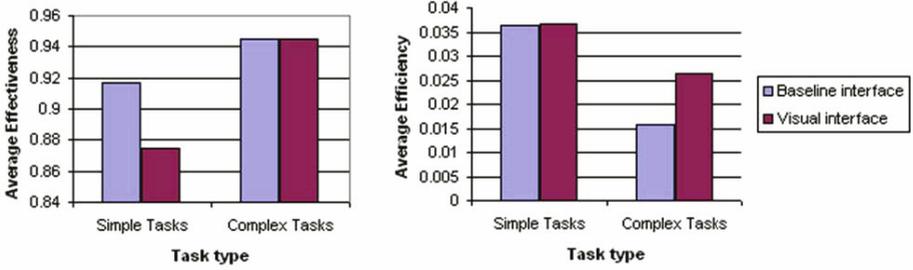
Fig. 7. Average Effectiveness and Temporal Efficiency by Task Type and Interface Used

of IVEA's design and development, the outcomes of the conducted study provide us with initial suggestive indications of its performances as well as potential usability problems.

With respect to the participants' performances, the average effectiveness and temporal efficiency are shown in Fig. 7.

In terms of effectiveness, it is interesting to note that, on average, the participants achieved a better score on the set of simple tasks by using the baseline interface than by using IVEA. We believe that this could be due to the effect of experience, whereby all the participants were very familiar with using keyword-based search interface to find out simple facts but had no prior experience with IVEA before. Meanwhile, they achieved the same score on the set of complex tasks by using the two interfaces, and the score was higher than those achieved on the set of simple tasks. This could be attributed to the order in which the task types were introduced (from simple to complex) whereby the subjects learned from the mistakes they made when using both interfaces for the first time while performing the simple tasks, and they became more experienced with both interfaces' features as they moved along to carry out the complex tasks.

When the task completion time was taken into consideration, Fig. 7 shows that the temporal efficiencies of the participants are comparable on the set of simple tasks by using the two interfaces. Meanwhile, on the set of complex tasks, the participant were considerably more efficient by using IVEA than by using the baseline interface. Moreover, the overall temporal efficiency of the participants on the set of complex tasks is lower than that of the participants on the set of simple tasks on both of the interfaces. The results suggest that our work is toward the right direction, however we are aware that a further usability study on a large scale is necessary.

With respect to the participants' satisfaction, their responses on the questionnaire are shown on Fig. 8. Among the 6 participants, a particularly well-liked aspect of IVEA was that its design and layout were appealing, whereas the ratings were a bit low for its self-descriptiveness and being complex for the task. We believe that with proper documentation, these setbacks can be mitigated. Some of the participants suggested to display 10 frequent phrases instead of the default
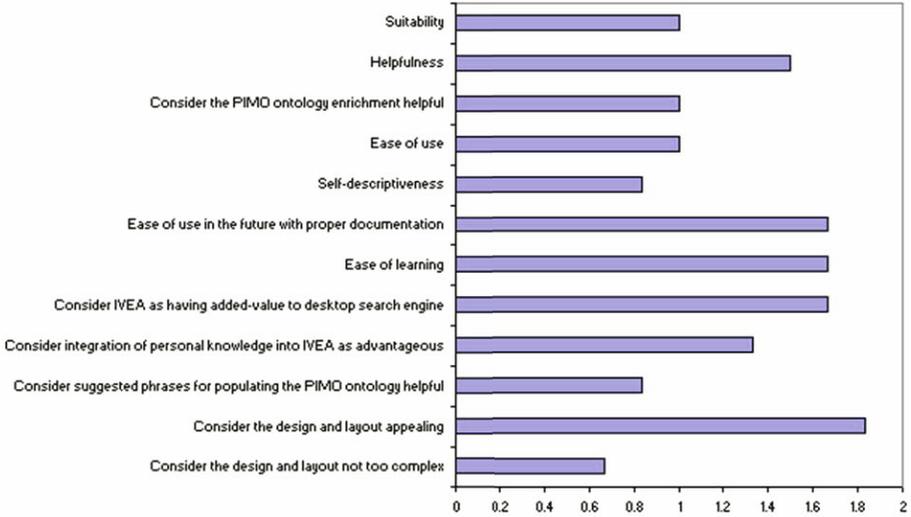
Fig. 8. Questionnaire responses

setting of 5. In fact, this number could be customized, but the lack of detailed manual rendered the participants uninformed of this feature. The participants also provided us with valuable design feedback such as:

- Highlight the most relevant document on the scatter plot.
- Group the visual items in the document view and the Entities Distribution View by class.
- Display the text surrounding the PIMO instances in the tooltip text of the TileBars' cells.
- Enable the Collection Overview to display at more than 2 dimensions.

To sum up, the study provided us with useful suggestive indications and valuable feedback to improve the design of IVEA in the future. In the next section, we review related work.

5   Related Work

Many research efforts in the area of Semantic Desktop [15,16,17,18] have introduced semantically-enabled Personal Information Management environments on the desktop. Although many useful functionalities are provided in these environments, we are not aware of any visualization tool that enables the users to explore and analyze a text collection according to their interests.

In the Information Visualization literature, a number of exploratory visualization tools are described in [19]. These tools, however, visualize different aspects of a domain based on the contents of *a single file*, which contains manually

cleaned metadata describing a set of documents within that domain [19]. They do not enable exploring a text collection based on the documents' contents, and the considerable efforts required to produce the metadata file make it impractical for wide-spread use. In addition, several document corpus visualization tools have been developed (e.g. [1,2,3,4,20]). However, they are based only on the textual contents of the documents and provide no mechanism to integrate and utilize an ontology representing the users' knowledge or interests. Hence, they lack some of the capabilities that ontology-based visualization tools offer, e.g. they do not allow the users to personalize the dimensions in which insights are to be gained. Nor do they enable the addition of newly discovered entities into the users' knowledge. It is also worth noting Aduna's AutoFocus, an ontology-based visualization tool which displays search results for documents on the desktop as clusters of populated concepts [21]. While allowing for the exploration of a document collection based on the extracted metadata and search keywords, in AutoFocus no deep insights or comparisons can be attained when the degrees of relevance of documents with respect to each entity within the users' spheres of interests are not available. Many other ontology-based visualization tools are designed specifically for certain domains and hence have different objectives than ours.

Also related to our work are faceted search and browsing tools, e.g. Flamenco [22] or Exhibit [23], which allow the users to navigate along different conceptual dimensions of a given document collection or a structured dataset. Although these tools are very effective in supporting browsing and exploratory tasks, they were designed either to work on document collection whose rich metadata categories are available or only to be used for browsing structured data.

6   Conclusions and Future Work

In this paper, we introduced IVEA, an innovative information visualization tool which supports the task of personalized exploratory document collection analysis. IVEA leverages upon the PIMO ontology as a formal representation of the users' interests and combined various visual forms in an innovative way into an interactive visualization tool. The integration of the PIMO ontology into IVEA allows the users to rapidly explore text collections at different levels of detail. To cater for the knowledge workers' evolving spheres of interests, IVEA enables them to enrich their PIMO ontologies with new concepts and instances on-the-fly. A small-scale usability study was carried out and the results were sufficiently encouraging to make it worthwhile to conduct a larger-scale usability study.

In future work, we plan to improve IVEA's functionalities based on the useful feedback gathered. In addition, we intend to employ anaphora resolution techniques so that co-references of PIMO instances (e.g. "European Central Bank", "ECB", "the bank", "it") can be correctly identified. Furthermore, the users should be able to tag documents with PIMO concepts and instances. IVEA will also need to be capable of handling more types of textual documents than just PDF files. Once the final version of IVEA is developed, a full-scale summative evaluation is to be carried out to gauge its usability.

As a final remark, it is our hope that the benefits that IVEA can bring about are not confined within its targeted task. If its final version has a high usability, it will demonstrate, to a certain extent, that the PIMO ontology is very useful and hence will motivate more users to start populating their own PIMO ontologies and use them to describe desktop items. This activity is certainly of particular importance to the realization of Semantic Desktop environments.

References

1. Shaw, C.D., Kukla, J.M., Soboroff, I., Ebert, D.S., Nicholas, C.K., Zwa, A., Miller, E.L., Roberts, D.A.: Interactive Volumetric Information Visualization for Document Corpus Management. Int. J. on Digital Libraries 2, 144–156 (1999)
2. Zhu, W., Chen, C.: Storylines: Visual exploration and analysis in latent semantic spaces. Computers & Graphics 31, 338–349 (2007)
3. Fortuna, B., Grobelnik, M., Mladenic, D.: Visualization of Text Document Corpus. Informatica 29, 497–504 (2005)
4. Grobelnik, M., Mladenic, D.: Visualization of News Articles. Informatica 28, 32 (2004)
5. Sauermann, L., Bernardi, A., Dengel, A.: Overview and Outlook on the Semantic Desktop. In: Decker, S., Park, J., Quan, D., Sauermann, L. (eds.) Proc. of Semantic Desktop Workshop at the ISWC, Galway, Ireland, November 6, 2005, vol. 175 (2005)
6. Sauermann, L., van Elst, L., Dengel, A.: PIMO - a Framework for Representing Personal Information Models. In: Pellegrini, T., Schaffert, S. (eds.) Proc. of I-Semantics 2007, JUCS, pp. 270–277 (2007)
7. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: IEEE Visual Languages, pp. 336–343 (1996)
8. Roberts, J.C.: State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In: CMV 2007: Proceedings of the 5th Int'l Conf. on Coordinated and Multiple Views in Exploratory Visualization, pp. 61–71. IEEE Computer Society, Washington, DC, USA (2007)
9. Hearst, M.A.: TileBars: visualization of term distribution information in full text information access. In: CHI 1995: Proc. of the SIGCHI Conf. on Human factors in computing systems, pp. 59–66. ACM Press/Addison-Wesley Publishing Co., New York (1995)
10. Reiterer, H., Mussler, G., Mann, T., Handschuh, S.: INSYDER - An Information Assistant for Business Intelligence. In: SIGIR 2000: Proc. of the 23rd Annual Int'l ACM SIGIR 2000 Conf. on Research and Development in Information Retrieval, pp. 112–119. ACM press, New York (2000)

11. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (2002)
12. Hatcher, E., Gospodnetic, O.: Lucene in Action. Manning Publications Co., Greenwich (2004)
13. Heer, J., Card, S.K., Landay, J.A.: prefuse: a toolkit for interactive information visualization. In: CHI 2005: Proc. of the SIGCHI Conf. on Human factors in computing systems, pp. 421–430. ACM Press, New York (2005)
14. ISO: ISO9241-11:1998: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability (1998)
15. Sauermann, L., Grimnes, G.A., Kiesel, M., Fluit, C., Maus, H., Heim, D., Nadeem, D., Horak, B., Dengel, A.: Semantic Desktop 2.0: The Gnowsis Experience. In: Proc. of the ISWC Conf., pp. 887–900 (2006)
16. Karger, D.R., Bakshi, K., Huynh, D., Quan, D., Sinha, V.: Haystack: A General-Purpose Information Management Tool for End Users Based on Semistructured Data. In: CIDR, pp. 13–26 (2005)
17. Cheyer, A., Park, J., Giuli, R.: IRIS: Integrate. Relate. Infer. Share. In: Decker, S., Park, J., Quan, D., Sauermann, L. (eds.) Proc. of Semantic Desktop Workshop at the ISWC, Galway, Ireland, November 6, 2005, vol. 175 (2005)
18. Groza, T., Handschuh, S., Moeller, K., Grimnes, G., Sauermann, L., Minack, E., Mesnage, C., Jazayeri, M., Reif, G., Gudjonsdottir, R.: The NEPOMUK Project – On the way to the Social Semantic Desktop. In: Proc. of the 3rd International Conference on Semantic Technologies (I-SEMANTICS 2007), Graz, Austria (2007)
19. Plaisant, C., Fekete, J.D., Grinstein, G.: Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository. IEEE Transactions on Visualization and Computer Graphics 14, 120–134 (2008)
20. Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., Williams, J.G.: Visualization of a document collection: the VIBE system. Information Processing and Management 29, 69–81 (1993)
21. Fluit, C., Sabou, M., van Harmelen, F.: Ontology-based Information Visualisation: Towards Semantic Web Applications. In: Geroimenko, V. (ed.) Visualising the Semantic Web, 2nd edn., Springer, Heidelberg (2005)
22. Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI 2003: Proc. of the SIGCHI Conf. on Human factors in computing systems, pp. 401–408. ACM, New York (2003)
23. Huynh, D., Karger, D., Miller, R.: Exhibit: Lightweight Structured Data Publishing. In: 16th Int'l WWW Conf., Banff, Alberta, Canada, ACM, New York (2007)