Designing a View for Visually Representing Information Coherence in a Document set

Louis Engelbrecht
CSIR, University of South Africa (UNISA)
South Africa

Dr. Adele Botha
CSIR, University of South Africa (UNISA)
South Africa

Abstract
The majority of the world's information is contained in textual records that describe specific domain information. The truth value of the information contained in these records is open for scrutiny. During the development of the National Indigenous Knowledge Management System (NIKMAS) system to capture and preserve indigenous knowledge, it was discovered that there was sometimes a coherence between the descriptions of the knowledge. It is proposed by the paper that a visual representation of the coherence of indigenous knowledge in a selected knowledge set would enable the user to make a decision about the possible truth value of the set. The aim of this paper is therefore to present an interface design for visually representing the coherence of information in a document set. In order to ensure that a rigorous process is followed, the design science research paradigm is suggested as appropriate research approach. In addition to interface design, a process for demonstrating and evaluating the design is also provided. The design presented in this paper assumes that the size of the document set is limited to not more than 50 documents.

Keywords
Information visualisation, information visualisation design, document visualisation, indigenous knowledge.

Introduction

A vast amount of information is stored in unstructured, textual records. The contents of these records usually describe facts about a specific domain in a natural language like English (McCallum 2005). Indigenous knowledge is such a domain where facts are stored in textual records. These facts describe information about indigenous knowledge such as traditional medicine and traditional food.
As part of an initiative by the South African Department of Science and Technology (DST) to capture and preserve indigenous knowledge, the National Indigenous Knowledge

Steyn, J., Van Belle, J.P. (Eds.). 2015. Beyond development. Time for a new ICT4D paradigm?
Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar: 177-188
ISBN: 978-0-620-68395-1

Management System (NIKMAS) was developed (Fogwill, Viviers, Engelbrecht, Krause & Alberts, 2011). During the development of the system it soon became apparent that there was some coherence observed between certain factual elements of certain traditional medicine descriptions.

The Coherence theory of truth, according to Young (2013) states that “the truth of any (true) proposition consists in its coherence with some specified set of propositions”. Roche (2013) in addition postulates that the coherence of a proposition with other propositions indicates an *increase* in the probability that it contains some truth. The ability to make a decision about the coherence of factual information contained in a number of indigenous knowledge records, therefore provides an insight into the possible truth value of the set. The design of an artifact to visually represent the coherence of facts in a document set is the focus of the paper and forms part of a larger research study to develop a model for visually representing the coherence of information in a document set. Engelbrecht et al. (2015) propose that design science research is an appropriate research paradigm for conducting visualisation research, and is therefore deemed appropriate for the larger study.

Hevner and Chatterjee (2010) indicate that design science research consists as 3 research cycles as depicted in *Figure 1*.

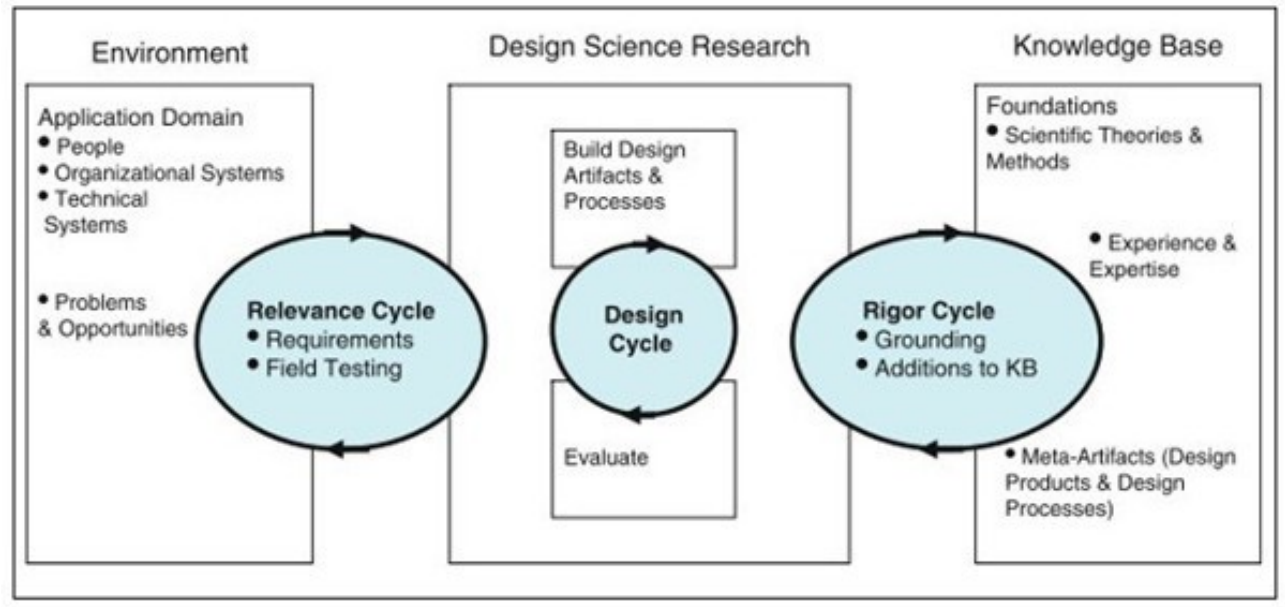

Figure 1. Design science research framework (Hevner & Chatterjee, 2010)

According to Hevner and Chatterjee (2010) the *Relevance cycle* aligns the context of the research project with the activities of the research, the *Design cycle* consists of artifact building and artifact evaluation activities and the *Rigor cycle* provides a scientific knowledge base to the research activities.

The *Rigor cycle* of design science research requires that the research activities should be informed by a scientific knowledge base. Visualisation design principles and guidelines defined by Engelbrecht et al. (2015) forms part of the knowledge base, and was therefore considered during the design of the interface.

The remainder of this paper is laid out as follows. In the following section the summarised list of design principles and guidelines provided by Engelbrecht et al. (2015) is briefly discussed. The discussion is followed by a description of the design of an interface to visually represent the coherence of information contained in a document set. The design of the interface is followed next by a discussion of a possible database schema design to support the interface. A brief discussion on the demonstration and evaluation of the visual representation is thereafter articulated. The paper is concluded in the final section.

Information visualisation design principles and guidelines provided by Engelbrecht et al. (2015)

The summarised list of design principles and guidelines provided by Engelbrecht et al. (2015) is briefly discussed in this section. Engelbrecht et al. (2015) postulate that the principles and guidelines provided by them should be considered during the design of a visualisation interface.

The design and development of Information System artifacts is considered by Peffers et al. (2006) as one of the activities of design science research. According to Hevner and Chatterjee (2010), the *Rigor cycle* of the design science framework, provides the scientific knowledge base to the research activities. The activity of designing a user interface for a visualisation could be informed by a scientific knowledge base in the form of a set of design principles and guidelines. Such a set is provided by Engelbrecht et al. (2015).

Engelbrecht et al. (2015) propose that their list of 31 information visualisation design principles and guidelines can be grouped according to the following concerns:

- User interactions and views.
- Additional information required by the user of the visualisation.
- The profile of the user.
- The Eppler and Burkhard Knowledge Visualisation framework.
- Forsell and Johansson's heuristic set for Information Visualisation evaluation.
- Elmqvist and Fekete's guidelines for implementing visual hierarchical aggregates.
- Visualisation shortcomings.

It is acknowledged by the research study that authors other than Engelbrecht et al. (2015) also compiled principles and guidelines that could inform the design of the information visualisation artifact. The design principles and guidelines provided by Engelbrecht et al. (2015) is proposed as appropriate since it summarises principles, guidelines and views provided by a number of other authors including (Ku, Nguyen, & Leroy, 2012), (Eppler & Burkhard, 2007), (Forsell & Johansson, 2010), (Elmqvist & Fekete, 2010), (Dix, 2013), (Shneiderman, 1996), (Chen et al., 2009), (Bihanic & Polacsek, 2012) and (Keller & Tergan, 2005).

A design for visually representing the coherence of information in a document set is articulated in the next section.

A design for the visual representation of the coherence of information in a record set

The design for visually representing the coherence of information in a document set is discussed in this section. Based on Shneiderman's (1996) information seeking tasks,

Engelbrecht et al. (2015) adds to their visualisation design principles and guidelines that the user should be presented with an information overview interface as well as the ability to zoom in on specific items. Based on this, the visualisation design presented by the research paper in essence consists of three different interfaces, namely

1. An *overall* view interface.
2. A *zoom-into-document* interface.
3. A *zoom-into-fact* interface.

The remainder of this section discusses the design of the *overall view* interface*,* the *zoom-into-document* interface and the *zoom- into-fact* interface.

Design of the *Overall View* Interface

The *overall view* interface is the first page presented to the user when he selects to enter the visualisation. The page visually represents an overall view of the document set by displaying the documents as well as the factual information in the set.
The simplified design of the *overall view* interface is depicted in *Figure 2*. The interface is divided into the following 5 sections:

- A *Visualisation* section (indicated by bullet 1).
- A *Browsing-and- filtering* section (indicated by bullet 6).
- A *Domain information* section (indicated by bullet 8).
- A *Document information* section (indicated by bullet 9)
- A *Document set information* section (indicated by bullet 10)

The decision to divide the interface into the specific sections, was based on the following principles and guidelines provided by Engelbrecht et al (2015):

- Additional information required by the user should be provided.
- Consider providing brushing-and-linking user interactions.
- Consider providing an overview.
- Consider providing functionality to filter out items.

The decision to specifically provide the *domain information section*, *document information section* and *document set information* sections was based on Chen (2005) and Chen et al. (2009) which states that in order for the user to make sense of the visualisation, it is required that information about the domain described by the data, as well as information about the data itself should be provided.

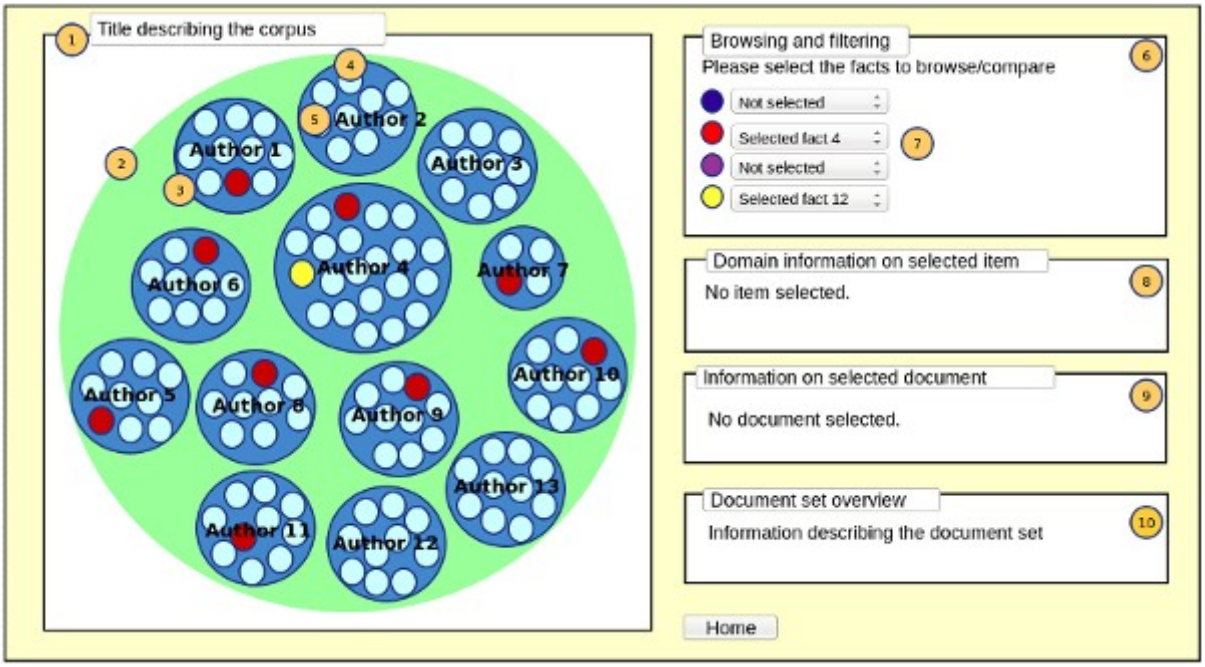

Figure 2. Design of the overall view interface

Each circle in the view is considered to be a visual hierarchical aggregate as it visually aggregates the underlying information on multiple levels.

- *Level 1*, indicated by bullet 2, is represented by the largest green circle and visually represents the document set as a whole.
- *Level 2*, indicated by bullet 3, is represented by the smaller dark-blue circles inside the large green circle. Each circle on this level represents a document in the set.
- *Level 3*, indicated by bullet 4, is the lowest level aggregate and contains facts in a specific document. Each fact is represented by the smallest circles in the view.

The design of the *overall view* interface is discussed next by considering each bullet item in *Figure 2*:

1. A title of the document set.
2. The large circle represents the whole document set. Each document in the set is represented by smaller circles inside this circle.
3. Each document in the set is represented by a circle, each containing a smaller circle that represents facts in the document. Clicking on any of the document representing circles zooms into the document, presenting the user with the *zoom-into-document* interface as depicted in *Figure 3*.
4. Each fact in the document is represented by the smallest circles. The bigger circle represents the document containing the fact. Clicking on a small circle, zooms into the information item, presenting the user with the *zoom- into-fact* interface as depicted in *Figure 4*.
5. The name of the author of each document is indicated, keeping display size limitations in mind.
6. The *Browsing-and-filtering* section allows for the user to perform Browsing and filtering actions on the document set. The section contains four dropdowns lists.
7. By selecting any of these dropdowns, the user selects to be presented with a view that indicates in which documents the fact occurs.

Steyn, J., Van Belle, J.P. (Eds.). 2015. Beyond development. Time for a new ICT4D paradigm?
Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar: 177-188
ISBN: 978-0-620-68395-1

8. The *Domain information section* displays domain information for a fact selected in the *visualisation section.*
9. The *Document information section* displays information for the document selected in the *visualisation section.*
10. The *Document set information* section provides the user with additional background information of the source of the document set.

The example provided in *Figure 2* illustrates how the coherence of facts in a document set could be visually represented. All the occurrences of a selected fact (Fact 4), can be indicated by coloring (red) in the circles representing the fact. The user can see which documents contain the same fact (Fact 4) and can furthermore make a decision on the coherence of that fact in the documents. It is proposed that having the knowledge of whether the facts are in coherence, allows for the user to acquire an insight into the validity of the facts and the document set as a whole.
In addition it is also possible to select more than one dropdown in section 6, allowing for the user to also contrast different facts (Fact 4 versus Fact 12) or to combine facts to acquire new insights.
After having discussed the design of the *overall view* interface, the next section considers the design of the *zoom-into-document* interface.

Design of the *Zoom-Into-Document* Interface

Selecting a document in the visualisation section presented in *Figure 2*, results in an animated, zoom-into effect which ultimately presents the user with the *zoom-into-document* interface as indicated in Figure 3.

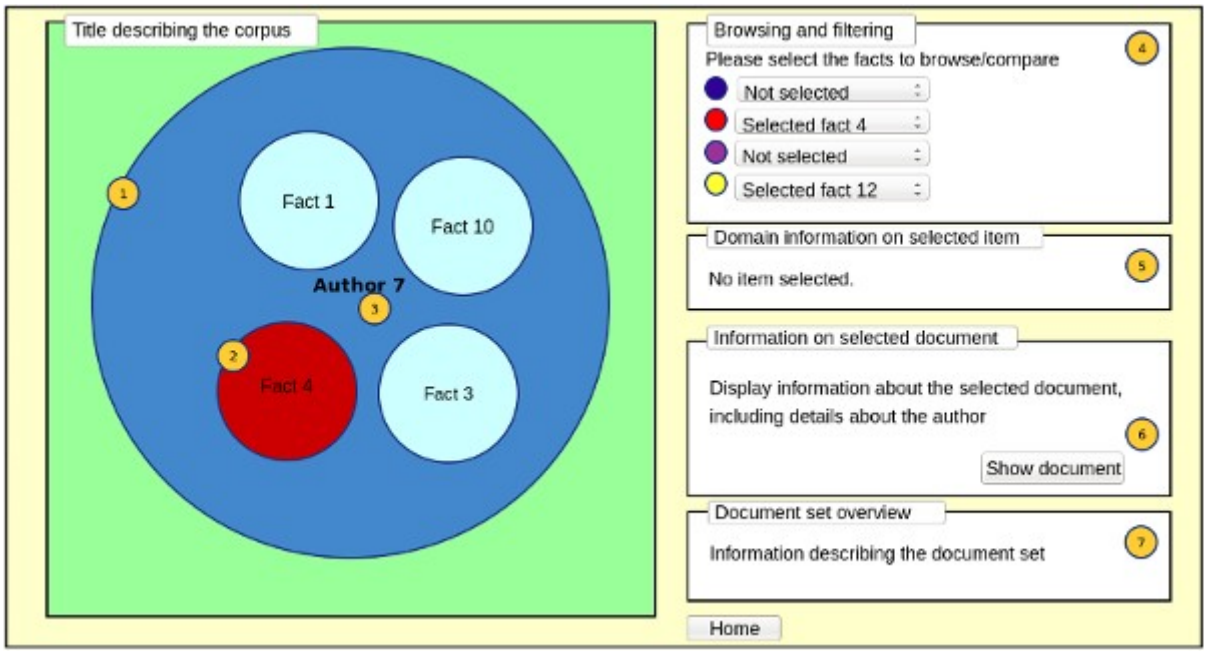

Figure 3. Design of the zoom-into-document interface

The design of the *zoom-into-document* interface is discussed next by considering each

bullet in *Figure 3*:

1. The large circle presented on this view, represents the document which was zoomed into from the *overall view* interface. Clicking on the area outside the circle zooms back to the *overall view*.
2. Facts contained in the zoomed into document are represented by smaller circles inside the large circle. Facts which were previously selected for a visual indication of their coherence with other facts, are still filled in with the indicated colour. Clicking on the fact representing circle, zooms into the fact, ultimately presenting the user with the view in *Figure 4*.
3. The name of the author of the selected document is displayed.
4. Facts selected for Browsing and filtering are not changed when zooming into the document.
5. No domain information is displayed since no facts are selected yet.
6. Information about the author of the zoomed into document is displayed. The information includes the name of the author, the author's date of birth and date of death, the author's location when writing the document, the name of the author's spouse, the size of the document and the date when the document was written. The aim of the information is to provide the user with background information about the source of the document; therefore enabling him to make an assumption about the quality of the information contained in the document. The information of the author is populated by a separate, manual process. Details of the process do not form part of this research paper. The text of the document is displayed in a new window by clicking on the *Show document* button.
7. Additional background information on the source of the document set is provided to the user.

After having discussed the design of the *zoom-into-document* interface, the next section articulates the design of the *zoom-into-fact* interface.

Design of the *Zoom-Into-Fact* Interface

Selecting a fact to zoom into, results in an animated, zoom-into effect which ultimately presents the user with the *zoom-into-fact* interface as indicated in Figure 4.
The design of the *zoom-into-fact* interface is discussed next by considering each bullet:

1. The circle presented in *Figure 4* represents the fact zoomed into from either the *overall view* interface (depicted in *Figure 2*) or the *zoom-into-document* interface (depicted in *Figure 3*). Clicking outside the circle returns the user to the preceding interface.
2. Known or researched domain information on the fact is presented to the user as part of the zoom-into action. The domain information is populated by a separate manual process, which is considered as being outside the scope of this paper.
3. Clicking on the *Show in context* button, opens the textual document in a new window, with all occurrences of the fact highlighted.
4. Zooming into a fact also zooms into the containing document. The information about the document is also displayed to the user.
5. The text of the document is displayed by clicking on the *Show document* button.
6. Additional background information on the source of the document set is provided to the user.

Steyn, J., Van Belle, J.P. (Eds.). 2015. Beyond development. Time for a new ICT4D paradigm?
Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar: 177-188
ISBN: 978-0-620-68395-1

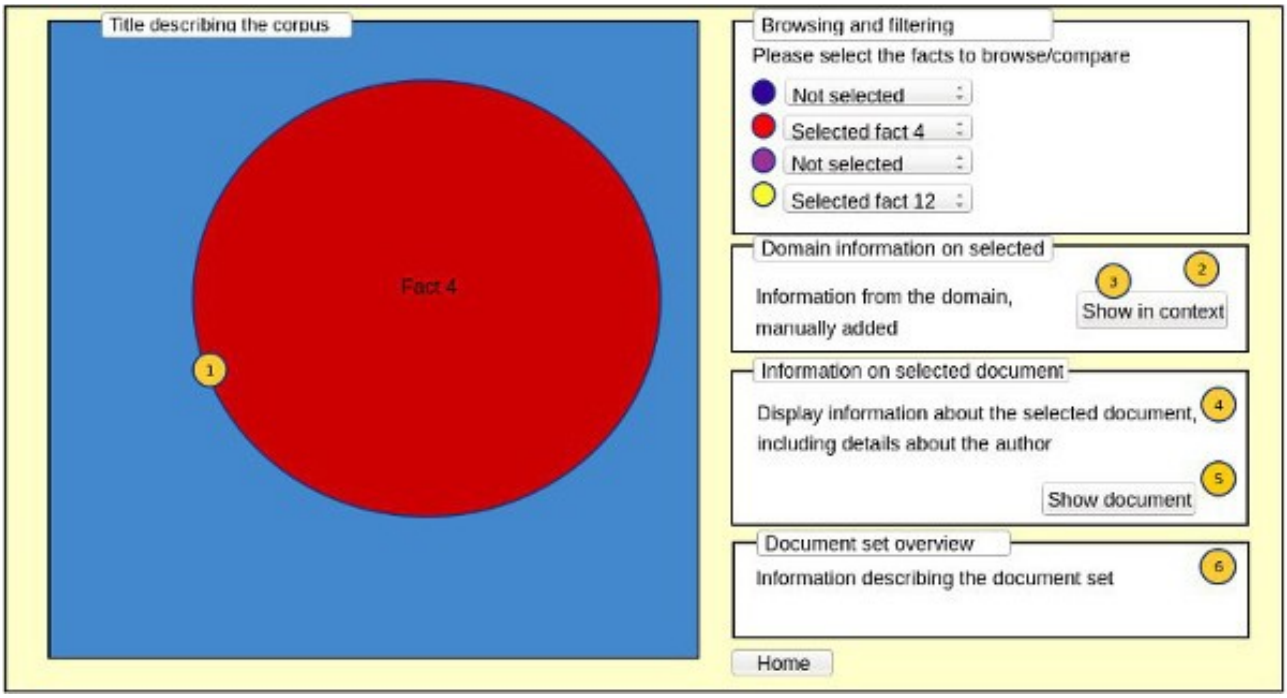

Figure 4. Design of the zoom-into-fact interface

Having articulated the design of the *overall view*, the *zoom-into-document* and the *zoom-into-fact* interfaces, the next section provides a possible database schema design for the views.

Database schema design for the proposed information visualisation solution

A database schema design to support the design of the visualisation interfaces of the study, is discussed next. The database schema design is depicted in *Figure 5* and consists of the following 10 tables:

- *Corpus*. Stores background information on the document set.
- *Document*. Stores background information on each document in the set.
- *Corpus_document*. Links each document to the *Corpus* table, specifying the documents belonging to the set.
- *Factindocument*. Each fact contained in a specific document is stored in this table.
- *Fact*. Stores information describing each fact contained in the document set.
- D*ocument_factindocument* Links facts to the documents in which they occur.
- *Factoccurence*. Stores information consisting of the start index and the end index of the occurrences of a fact in a document.
- *Factindocument_factoccurrence*. Links each fact to the *Factoccurence* table, indicating all the occurrences of a fact in each document.
- *Information*. Stores the facts contained in the document set.

- *Domaininformation*. The table describes information from the relevant subject domain on each fact contained in the documents.

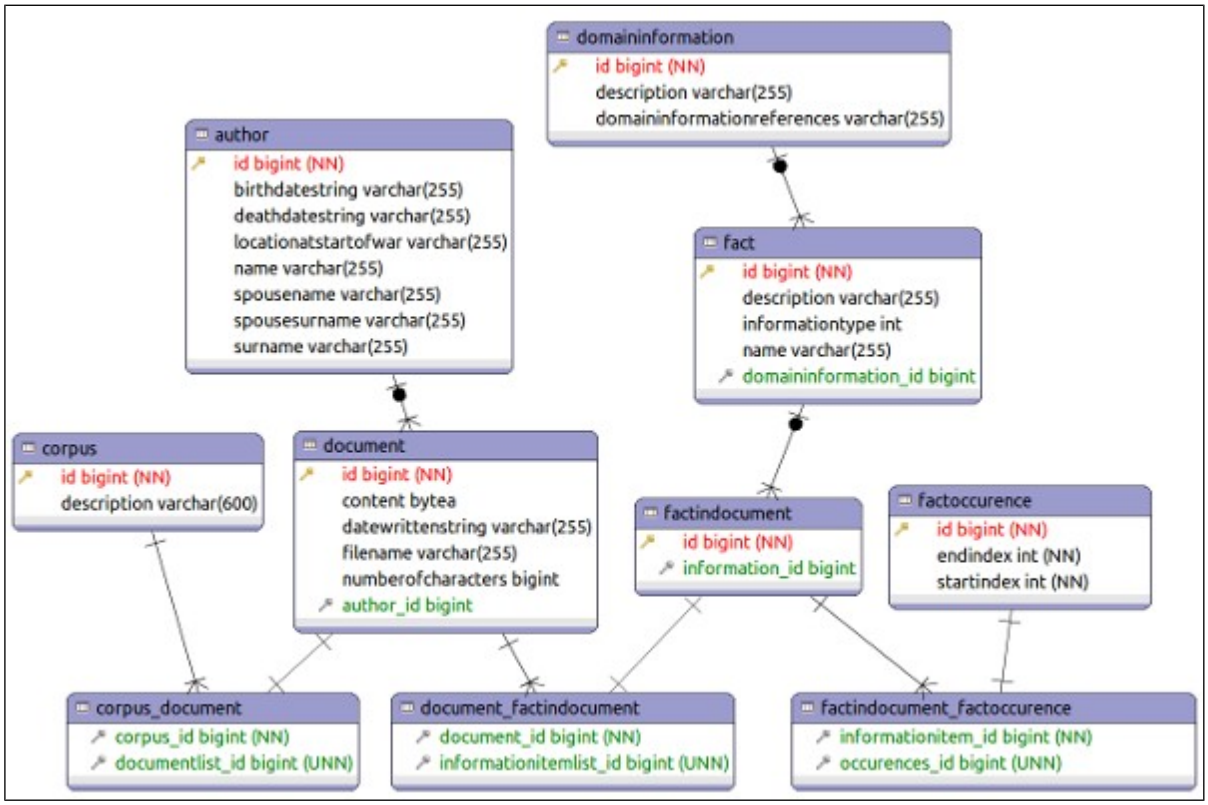

Figure 5. Database design

Having briefly discussed the design of the database schema to support the user interface, the next chapter describes how the design is to be demonstrated and evaluated.

Demonstration and evaluation

A discussion on how the proposed visualisation design could be demonstrated and evaluated, is discussed in this section. Peffers et al. (2006) propose that design science research can be conducted by executing the following six activities:

1. *Problem identification and motivation.* Includes the identification of the research problem as well as the motivation for finding a solution to the problem.
2. *Objectives of a solution.* This activity in the research process is concerned with the definition of the objectives of the solution. According to Peffers et al. (2006), these objectives can be either qualitative or quantitative and must be inferred from the problem definition.
3. *Design and development.* Involves the designing and building of the Information Systems artifact. This research paper focus mainly on the design activity.
4. *Demonstration.* The use of the artifact to solve the problem is demonstrated during the activity.
5. *Evaluation*. Involves an observation of how well the artifact supports the objectives of the solution as set out in activity 2.

6. *Communication*. The results of the research are finally communicated through a variety of means, including scholar articles.

In order to demonstrate and evaluate the interface design, it has to be transformed into a workable software application. This paper is only concerned with the interface design; the process of extracting information from the indigenous knowledge descriptions does not form part of this article. After building the Information Systems artifact, it can be demonstrated to a number of research participants. It is suggested that for the purpose of the study, the resulting application can be demonstrated to a number of 5 indigenous knowledge domain experts as well as 5 IT experts. The decision to use 5 domain experts and 5 IT experts is based upon Nielsen and Landauer's (1993) proposition that between 3 and 5 expert users is adequate for conducting user evaluations.
It is proposed that during the demonstration activity, a short feedback questionnaire can be presented to the audience. The feedback received from the questionnaire can subsequently be used as input to the evaluation activity. The questions in the feedback questionnaire can be in the form of Likert scale statements. For each statement the participant has to indicate how strongly he/she agrees or disagrees by selecting a value from 1 to 5 according to a 5-level Likert scale. The following statements as suggested by Engelbrecht et al (2015) can be included in the questionnaire:

1. The technique for visualising the facts contained in the records is appropriate.
2. The visualisation technique employs a minimal set of actions to accomplish a task.
3. Functionality to control the level of details is available.
4. The visual layout of the visualisation is effective.
5. The cognitive load on the user is minimal.
6. The visualisation contains extracted information.
7. There is too much information on the interface.
8. Additional functionality to show less information should be considered.
9. The visualisation is clean and simple.
10. Data items are easily distinguishable from each other.
11. The visualisation can lead to misleading interpretations.
12. The visualisation is interpretable.
13. The user can view what he needs to see.
14. The visualisation can be used to get some insights into the domain described by the data.

The demonstrations can be followed by a statistical evaluation of the results in order to

- determine whether the interface is useful,
- is based on design principles and guidelines and
- whether it can be used to acquire an insight into the coherence of the facts in the document set.

The design for visually representing the coherence of information contained in a document set, as well as a proposal for the demonstration and evaluation of the design was articulated in the previous sections. A short conclusion is presented next.

Conclusion

As part of a larger research study to visually represent the coherence of facts in a textual document set, an interface design was presented in this paper. The facts can contain information about a domain like indigenous knowledge. Having a technique for visualising the coherence of facts in indigenous knowledge records can provide a mechanism for making an assumption on the truth value of indigenous knowledge.

Design science research was proposed as research approach for conducting the larger study. This paper forms part of the design and building activity of the design science research process; focusing on the design of the visual representation artifact. Suggestions for conducting the demonstration and evaluation activities of the process was also provided.

By visually representing the coherence of facts in indigenous knowledge and therefore enabling the user to make an assumption about the truth value of the facts, this paper contributes to the acknowledgement of indigenous knowledge in general as having value. Future work includes the development of an application based on the interface, demonstrating the application and evaluating its utility and effectiveness. Ultimately it is envisaged that the solution can in the future be applied to the NIKMAS system.
In addition to indigenous knowledge it is contemplated that the solution can also be applied to other fields of study where domain facts are described in records, including history (historical testimonies), law (court testimonies) and medicine (patient records).

References

Bihanic, D., & Polacsek, T. (2012). Models for Visualisation of Complex Information Systems. *2012 16th International Conference on Information Visualisation*, 130–135. doi:10.1109/IV.2012.32

Chen, C. (2005). Visualization Viewpoints: Top 10 Unsolved Information Visualization Problems. *Computer Graphics and Applications, IEEE*, (25.4), 12–16.

Chen, M., Ebert, D., Hagen, H., Robert, S., Shneiderman, B., & Tory, M. (2009). Visualization Viewpoints. Data, Information and Knowledge in Visualization. *Computer Graphics and Applications, IEEE*, (January/February 2009), 12–19.

Dix, A. (2013). Introduction to Information Visualisation. In *Information Retrieval Meets Information Visualization* (pp. 1–27). Berlin, Heidelberg: Springer.

Elmqvist, N., & Fekete, J.D. (2010). Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3), 439–54. doi:10.1109/TVCG.2009.84

Engelbrecht, L., Botha, A., & Alberts, R. (2015). Designing the Visualization of Information. *International Journal of Image and Graphics*, *15*(2), 1–19. doi:10.1142/S0219467815400057

Eppler, M. J., & Burkhard, R. A. (2007). Visual representations in knowledge management: framework and cases. *Journal of Knowledge Management*, *11*(4), 112–122.

doi:10.1108/13673270710762756

Fogwill, T., Viviers, I., Engelbrecht, L., Krause, C., & Alberts, R. (2011). A software architecture for an indigenous knowledge management system. In *Indigenous Knowledge Technology Conference (IKTC)*. Windhoek, Namibia.

Forsell, C., & Johansson, J. (2010). An heuristic set for evaluation in information visualization. In *International Conference on Advanced Visual Interfaces - AVI '10* (p. 199). New York, New York, USA: ACM Press. doi:10.1145/1842993.1843029

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems. Theory and Practise*. (R. Sharda & S. Voß, Eds.). Springer Science and Business Media. doi:10.1007/978-1-4419-5653-8

Keller, T., & Tergan, S. (2005). Visualizing Knowledge and Information: An Introduction. In T. Keller & S. Tergan (Eds.), *Knowledge and information visualization* (pp. 1–23). Springer Berlin Heidelberg.

Ku, C., Nguyen, J. H., & Leroy, G. (2012). TASC - Crime report visualization for investigative analysis: A case study. In *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)* (pp. 466–473). IEEE.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

McCallum, A. (2005). Information extraction: Distilling structured data from unstructured text. *ACM Queue*, *3.9*(November), 48–57.

Nielsen, J., Landauer T.K. (1993). A mathematical model of the finding of usability problems. In *INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 206–213). http://doi.org/10.1145/169059.169166

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The design science research process: A model for producing and presenting information systems research.

Roche, W. (2013). On the Truth-Conduciveness of Coherence. *Erkenntnis*, 1–19. doi:10.1007/s10670-013-9564-y

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, 336–343. doi:10.1109/VL.1996.545307

Young, J. O. (2013). The Coherence Theory of Truth. In *The Stanford Encyclopedia of Philosophy*. Retrieved from http://plato.stanford.edu/archives/sum2013/entries/truth-coherence/