Federating Resources of Information Systems: Browsing Interface

Andrei V. Malchanau, Paul E. van der Vet & Hans E. Roosendaal
Department of Computer Science, University of Twente
P.O.Box 217 7500AE Enschede, The Netherlands
{a.v.malchanau, p.e.vandervet, h.e.roosendaal}@utwente.nl

Abstract

Designing the user interface of a federated system (what we call a browsing interface) must consider the knowledge gap that exists between desires of the users and the needs the systems are built to support.

The concept of Habitable Interfaces aims to bridge the knowledge gap by providing kinds of representations and the interaction with these representations that are based on domain knowledge. Habitable Interfaces will allow the organising of currently disparate archives into cohesive domain specific federations of information resources.

To approach designing Habitable Interfaces we propose a model of communication and a criterion.

1. Introduction

There are varieties of information resources that are available for scientists through the Internet. These resources are heterogeneous such as databases and archives of documents and multimedia. More, there are resources that run algorithms rather than retrieving data. In many cases, these resources are built around a particular need of a local group of scientists that are collecting data (or writing algorithms) for a particular reason. Clearly, maintenance of this variety of information resources cannot be centralised. On the other hand, the scientists need an access to these resources regardless from the underlying technological differences. Federated systems aim at providing an access to and combining information from disparate and heterogeneous information resources.

There are several ongoing efforts worldwide aimed at designing federated systems as well as data warehouses.Some examples are [1, 4, 5, 6, 14]. Most of this work is being done on combining data and solving technological issues of creating federated systems. Although having resources readily accessible is a necessary condition, the user interface makes a difference between a collection of independent information resources and a federated system.

Van der Vet (2000) proposed a research environment to alleviate some of the issues of accessing web-based information resources.

As we noted above, there are many organisations maintaining information systems, and their number grows by the day (see, for example, an overview of information resources in molecular biology in [12]). Individual research groups generally will want to leave maintenance of these resources to the groups who created them. The organisation of the access to existing resources should better be based on federating these resources rather than on integrating them into monolithic systems [8].

When federating information resources a number of high level issues should be addressed:

- Interaction of the user with the system: retrieving data, interacting with these data and receiving user feedback
- Representing information about the content of a federated system and retrieved data to the user
- Combination of the data (a common schema and algorithms for combining data from different resources)
- Accepting, planning and optimising user's queries
- Communication with the resources

Building federated systems requires a design of the user interfaces that will allow users to utilise available information effectively and efficiently. Existing approaches to the design may not fit the scale of the federated systems.

Perhaps, the most important issue in creating a federated system, is the gap between a variety of possible views and classifications of the same facts and rules that constitute knowledge, on the one hand and the limited representation that a designer can show to the user on the other hand. In the following part, we consider some issues of this gap in more detail.

Interaction of the user with the system and representing information about the content of a federated system are the most relevant topics for this paper. This brings us to the concept of Habitable Interface.

2. The knowledge gap

Federating information resources brings up the issue of the gap between varieties of views on information stored in an archive and the necessarily limited design of the archive. Further, we argue that the gap is inherent in the communication process and calls for new approaches.

Scientists are engaged in a knowledge discovery process. Knowledge is accumulated by collecting data. Collecting data requires a model that serves the purpose of practical guidance. Knowledge discovery is a collective effort, and a collective effort needs communication. In communication, researchers have generally different roles of authors and readers [13]. Given the variety of purposes that knowledge can be applied to, and the variety of data models, it is next to inevitable that there is a mismatch between the reader's and the author's data model. The situation worsens, when there are many readers and many authors who are trying to communicate on similar issues.

The archive can be perceived as an intermediary between authors and readers. Building an archive requires yet another data model (Figure 1). Differences between an archive's data model and authors' data models are not an obstacle for communication as it is only a question of converting known data using known data models.
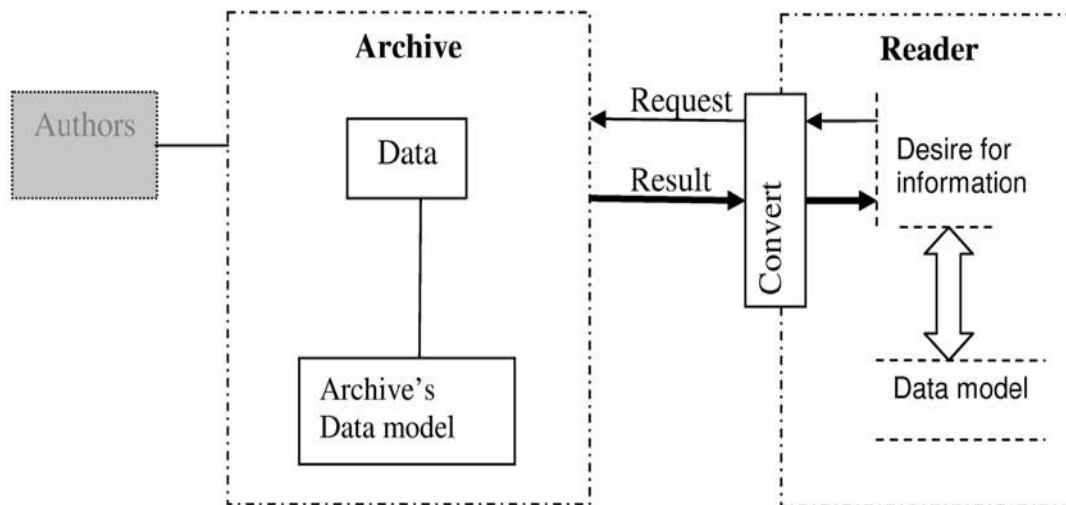

Figure 1. Communication between archive and reader

But for readers the situation is different. Readers do not need to know the archive data model and they do not want to know the archive data model, as it does not fit their mental frame. As a consequence, there is a gap between what we call desire (expressing what information the reader wants to know), on the one hand, and need (referring to the information in the archive's terms), on the other hand.

To fill the gap, an archive could convert data into a form required by readers. Multiplicity and dynamics of readers' interests present too great a challenge for designers of archives and in principle, even the best study of requirements would not provide a uniform representation of the readers' interests. Indeed, there is no average reader and there are many different archives.

3. Habitable Interfaces

Habitable Interfaces can help users to convert their desires for information into information needs that are then being communicated to the existing information resources. To arrive at an approach to designing Habitable Interfaces we start from a high-level model of communication between the reader and the archive. This model is rooted in other models proposed in the literature on Information Seeking and Information Retrieval: there are several overviews of the models and the concept of information in general (see for example [2, 10]). Here we would like to briefly consider the model proposed by R.S. Taylor [15] who describes the process of asking questions as starting from the 'visceral need':

- Visceral need – a vague dissatisfaction with the current knowledge about some topic;
- Conscious need – conscious understanding of what kind of information (knowledge) is missing;
- Formalised need – the formal statement of the need;
- Compromised need – the query that is presented to the archive

The model suggests that users formulate a query in several steps. In later experiments [3] the last three steps were reportedly observed. But Taylor's model does not explain how the visceral need is being converted into a compromised need. We believe that this conversion depends on the design of the retrieval system. An experimental investigation on such models has to generalize beyond the design of the system used in the investigation. In other words if the system design implies certain behaviour of the user it is likely to induce such a behaviour. For example with some interfaces, the readers have to explore the archive, with others they have to know the terms used in the archive before they can search for the desired information.

The significance of the Taylor's model for Habitable Interfaces is that it postulates that a request to the archive is a result of converting a particular 'inadequacy' in a reader's knowledge about some topic.

We add to this model of communication a model of the system. This allows stating a hypothesis about the system design that can be validated using empirical data.

To show how our model can be constructed we consider the communication process that takes place between the reader and an archive.

First, before starting the communication process, the reader has a certain desire for information. The word 'desire' implies a strong intention or aim. It is in contrast with the 'need' that is in general defined as a lack of something requisite or useful. Figure 2 shows this distinction from a number of viewpoints.
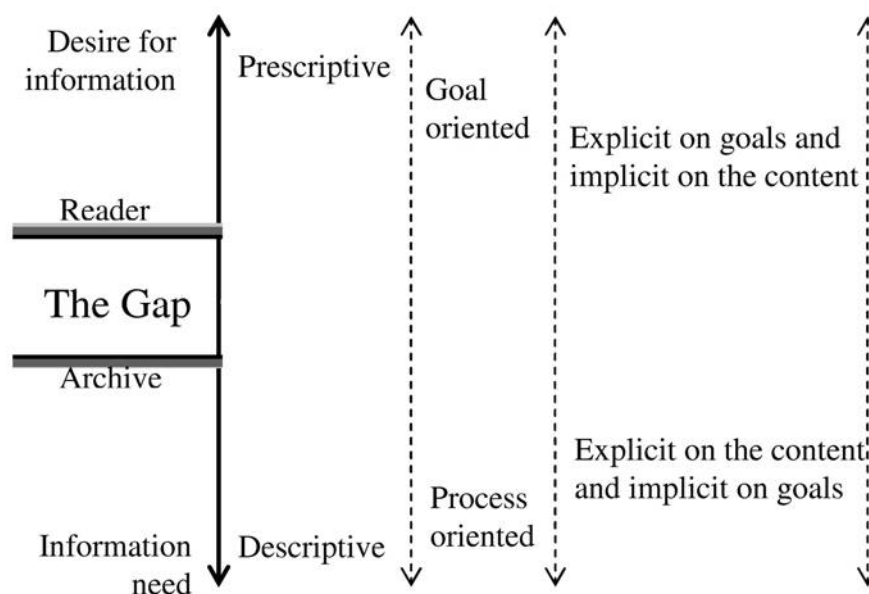

Figure 2. Distinction between desires and needs. There is a gap if the reader and the archive are at the different levels.

Based on the above we arrive at the following characteristics of the model of communication between reader and federated archive:

- The readers have knowledge that can be divided into classes serving the purpose of building a model.
  - Domain knowledge is a set of facts and rules that are known within a certain research domain. In a federated archive, this knowledge can form a basis to organise disparate resources.
  - We refer to the knowledge about the current situation as the situation perceived by the reader. It may be not an adequate understanding of the situation by the reader, but for our purposes, this is not relevant. Situations, in which the reader can be, may be rather divers. If we add to this an individual interpretation of the situation, it is clear that this knowledge will be specific for any individual reader. The knowledge about the current situation will set the context, within the domain, for the information being communicated.
  - The knowledge about the system or the language of the system is needed for converting the desire for information into a query that is comprehensible for the system and for converting results returned by the system into a form that the reader understands. The knowledge about the system, as any knowledge on communication, can be considered at four levels: lexical, syntactic, semantic and pragmatic. The fourth pragmatic level in this classification is the level at which the desire for information is communicated.
- The desire for information stems from the current situation and it is based on domain knowledge. The desire also indicates some lack of domain knowledge;
- Readers do not need to know the language of an archive and many readers will lack knowledge about the archive and its language;
- The gap between the desires for information of the readers and the information needs supported by archives can be viewed as the combined effect of lack of domain knowledge and lack of knowledge about archives;
- Converting this lack of knowledge into communication requests generates different sorts of behaviour, that, in general, depend also on the design of the archive;
- The communication process stops when the reader is satisfied.

This model is depicted in Figure 3. Figure 3 also shows the way the information resources can be organised into a federated archive. An important question to be answered based on this model is the design of the "Unravel" and "Combine" functions. "Unravel" function presents to the reader what is available in the federated archive and allows building a comprehensive set of queries. In the federated archive, the query has to be "Translated" into requests to individual resources, since the internal representation of the federated archive differs from that of an individual resource. The results returned by the resources might need translation, too. Furthermore, these results must be combined into a single representation for the reader. However, the particular implementation would require answers to questions such as:

- What are the principles to organise the representations of the information sent to and retrieved from the different resources?
- What sort of interaction between the representations in the "Unravel" and the "Combine" functions is useful and adequate for the reader?

Our model suggests that the representations should be based on the domain knowledge and the interaction with these representations should be designed so that it requires minimal knowledge of the system.

The design of these functions serves to reduce the requirements on the reader's knowledge about the system, and to improve the efficacy and efficiency of the communication.
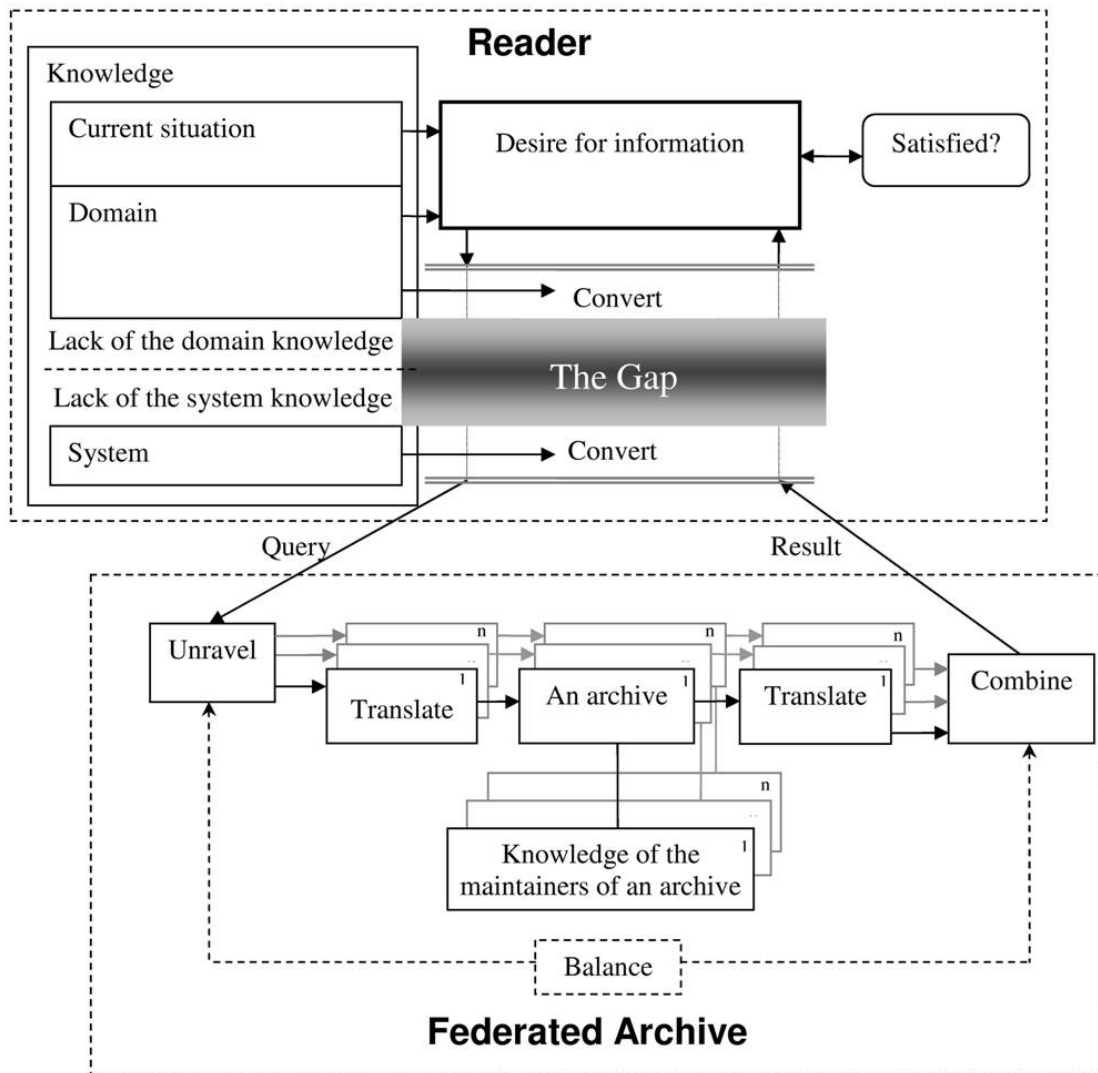
Figure 3. A model of Habitable Interfaces

3. Evaluation

For the high-level model to be applicable in designing a federated archive, it needs to be empirically validated. Such an evaluation can only be based on a priori agreed criteria and a method for evaluation. We argue that trust is a good indicator of the quality of the scientific communication that takes place and can be assessed using a mix of quantitative and qualitative methods.

In their communication, scientists are sharing with and delegating to the archive some of their tasks. In this perspective the reader, the trustor, should be able to trust the archive, the trustee, in this process of communication. This level of trust is posed to be a good indicator of the quality of information exchange. More on the relation of trust to scientific communication can be found in [9] and to information science and technology in [11].

4. Conclusions

Federating information resources requires new approaches towards designing user interfaces. The main issue is to deal with the degree of complexity and the scale of the integrated system.

On top of the complexity of the system there is a gap between the desires the users have and the needs that are supported by the systems.

The concept of Habitable Interfaces aims at helping the user to bridge this gap by means of incorporating the domain knowledge into representation and interaction.

At present, we work on the empirical validation of the model of Habitable Interfaces. The criteria of the validation will be based on the level of the user's trust in a federated system.

In addition, we would like to explore how intelligent agents may support the user in carrying out routine but specific tasks.

References

1. Blanco, J.M., Illarramendi, A., Goni, A. (1994) Building a Federated Relational Database System: An Approach Using a Knowledge-Based System, Int. Journal of Intelligent and Cooperative Information Systems, Vol.3, No.4 pp. 415-455
2. Capurro, R. & Hjørland, B. (2003) The concept of Information. In B. Cronin (Ed.), Annual review of Information Science and Technology, Vol. 37 pp. 343-411. Medford, New Jersey: Information Today
3. Chen, H. & Dhar, V. (1990) Online Query Refinement on Information Retrieval Systems: A Process Model of Searcher/System Interactions. In Proceedings of the 13th International Conference on Research and Development in Information Retrieval pp. 115-133. Brussels, Belgium, 5-7 September 1990
4. Chen, J., DeWitt, D., Tian, F. & Wang, Y. (2000) NiagaraCQ: A scalable continuous query system for internet databases. In Proc. of the ACM SIGMOD Conf. on Management of Data pp. 379-390
5. Davidson, S., Crabtree, J., Brunk, B., Schug, J., Tannen, V., Overton, C. & Stoeckert, C. (2001) K2/kleisli and GUS: Experiments in integrated access to genomic data sources. IBM Systems Journal, 40(2) pp. 512-531
6. Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajamaran, A., Sagiv, Y., Ullman, J., Vassalos V. & Widom, J. (1997) The TSIMMIS Approach to Mediation: Data Models and Languages, Journal of Intelligent Information Systems, 8 pp. 117-132
7. Gonçalves, M. A., France, R. K. & Fox E. A. (2001) MARIAN: Flexible Interoperability for Federated Digital Libraries. Research and Advanced Technology for Digital Libraries: Proceedings of the 5th European Conference, pp. 173-186. ECDL-01 (Darmstadt, Germany: 4-9 Sept.) Springer
8. Gray, P.M.D., Kemp, G.J.L. (2000) Federated database technology for data integration: lessons from bioinformatics. In Koslow, S.H., Huerta, M.F. Electronic collaboration in science pp. 45-72. Mahwah NJ: Lawrence Erlbaum
9. Hummels, H. & Roosendaal, H. E. (2001) Trust in Scientific Publishing, Journal of Business Ethics 34 pp. 87-100
10. Ingwersen, P. (1992) Information Retrieval Interaction. London: Taylor-Graham. X, 246 p.
11. Marsh, S., Dibben, M.R. (2003) The role of trust in Information Science and Technology. In B. Cronin (Ed.), Annual review of Information Science and Technology, Vol. 37 pp. 465-498. Medford, New Jersey: Information Today
12. Reed, J. (2000) Trends in Commercial Bioinformatics. Oscar Gruss
13. Roosendaal, H. E. & Geurts, P. A. Th. M. (1997) Forces and Functions in Scientific Communication: an Analysis of their Interplay, Proceedings of the Conference on Co-operative Research in Information Systems in Physics, University of Oldenburg, Germany, September 1-3
14. Stevens, R., Baker, P.G., Bechhofer, S., Ng, G., Jacoby, A., Paton, N.W., Goble, C.A. & Brass, A. (2000) TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources. Bioinformatics 16(2) pp. 184-186
15. Taylor, R.S. (1968) Question-negotiation and information seeking in libraries. College and Research Libraries, 29 pp. 178-194
16. Vet, van der, P.E. (2000) Building web resources for natural scientists, in: Interactive distributed multimedia systems and telecommunication services (Proceedings IDMS2000), H. Scholten and M.J. van Sinderen (eds.), Berlin: Springer, (LNCS 1905), pp. 205-210