

Cumulative Duality in Designing Information Brokers

B.C.M. Wondergem, P. van Bommel, Th.P. van der Weide,

Computing Science Institute/

CSI-R9808 March 1998

Computing Science Institute Nijmegen
Faculty of Mathematics and Informatics
Catholic University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands

Cumulative Duality in Designing Information Brokers

B.C.M. Wondergem, P. van Bommel, and Th.P. van der Weide
Computing Science Institute, University of Nijmegen
Toernooiveld 1, NL-6525 ED, Nijmegen
The Netherlands
(bernd@cs.kun.nl)

March 17, 1998

Technical Report CSI-R9808

Keywords: Information Discovery, Middle-Agents, Conceptual Agent Design, Information Retrieval, Multi-Agent Systems.

Abstract

The focus of this paper is information brokers within Information Discovery (ID). We describe Cumulative Duality matrices, an instrument to deal with design criteria for such information brokers.

ID is the synthesis of Information Retrieval and Information Filtering, where information brokers act as middle-agents. There are numerous design criteria for information brokers. Since these stem from ID, they exhibit a dual nature. The duality of the criteria is shown to be cumulative. In the form of a matrix, cumulative duality can be used as a design instrument for information brokers.

Contents

1	Introduction	3
2	Cumulative duality in Information Discovery	4
2.1	Duality in Information Discovery	4
2.2	Cumulative duality in directed communication	5
2.3	Cumulative Duality Matrix	5
3	Instantiations of CD Matrices	6
3.1	Partial environmental knowledge	6
3.2	Privacy of interest and content	8
3.3	Dynamical Interest and Content	9
4	Other Cumulative Dual Criteria	10
5	Conclusions & Further Research	11

1 Introduction

In the quest for relevant information, two conceptually *dual* approaches have emerged: Information Retrieval (IR) (e.g. [12]) and Information Filtering (IF) (e.g. [2]). Where IR deals with dynamic user needs and stable collections of information, IF is concerned with stable user needs and a dynamic supply of information. See [2] for a comprehensive comparison of IR and IF.

Our paradigm Information Discovery (ID), consisting of the synthesis of IR and IF, is more powerful than their mere combination ([18]). In ID, we consider a community of possibly interconnected users, having both stable and dynamic information needs. Also, we consider a number of sources, again stable and dynamic and possibly communicating, containing documents to satisfy users' information needs. This process is facilitated by intermediaries called *information brokers*, which cooperatively compare user interests with available documents. Users and sources communicate via brokers, as depicted by solid arcs in figure 1. Communication between entities of the same type is depicted by dashed arcs.

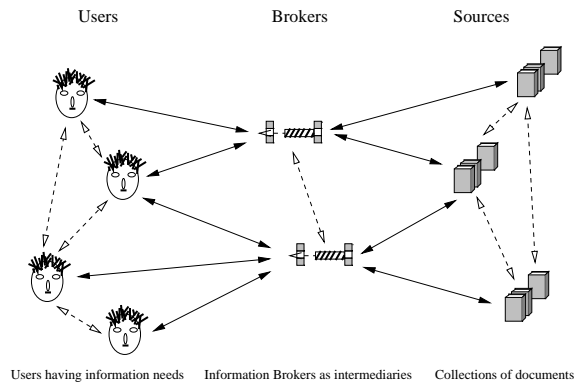


Figure 1: Information Discovery Paradigm

The paradigm in Figure 1 is augmented in several perspectives. Firstly, the entities in the ID paradigm are modeled as intelligent agents, since agent technology (see e.g. [19]) is perfectly suited for dynamic ID environments ([17]). Secondly, information brokers essentially are *middle-agents*. Middle-agents can be designed in a variety of ways (see e.g. [9] and [8]), resulting in different instantiations of the ID paradigm. We consider design criteria for information brokers regarding ID, concerning e.g. privacy and knowledge about locations of users and sources. These criteria exhibit a *dual* user-

source nature caused by duality in ID.

The focus of this paper is on the conceptual design of information brokers in the context of the ID paradigm. Our goal is to describe Cumulative Duality matrices, an instrument which explicitly shows the influence of dual ID criteria on the design of information brokers. As a consequence, this instrument provides a solid basis for analysis of duality in ID.

The duality of ID criteria is cumulative over the ID communication directions, i.e. either from user, via broker to source, or the other way around. Although the proposed application domain is ID, our approach is applicable in every paradigm that exhibits a dual nature and uses middle-agents. For instance, middle-agents in electronic commerce (see e.g. [15]) can be focused on. This also holds for numerous other applications of WWW agents, and middle-agents in general. As an example, consider BargainBot ([1]), helping users with shopping and IR, and Kasbah ([6]), aiding users in buying and selling goods.

The organisation of this paper is as follows. In section 2, duality of design criteria in ID is examined and Cumulative Duality matrices (CD matrices) are introduced. Section 3 provides different applications of CD matrices, leading to different broker designs. Section 4 elaborates on other design criteria for information brokers. Finally, section 5 provides concluding remarks and directions for future research.

2 Cumulative duality in Information Discovery

2.1 Duality in Information Discovery

Duality in ID stems from the synthesis of the dual paradigms of IR and IF. On a basic level, i.e. the level of contextual assumptions, this is evident in, for instance, dynamic vs. static user needs and sources of information. On a higher level, the goals of users and sources are dual: users want only relevant information to be rendered (their preference is to receive *not too much* information), whereas sources want to deliver as much information as possible (their preference is not to distribute *not too little* information).

A result of both forms of duality between users and sources is an imbalance in the ID paradigm. To guarantee fairness in an unbalanced paradigm, neutral intermediaries are introduced: information brokers. As stated in the introduction, the more basic goal of information brokers is the matching of user interests and available documents.

A common aspect of these information brokers and Negotiated Retrieval ([11]) is the merging of documents out of several sources. However, information brokers have more goals and features. Information brokers should, for instance, be proactive to relieve the user as much as possible.

2.2 Cumulative duality in directed communication

There is a clear relation with client-server architectures. Information brokers are true middle-agents, since all communication between users and sources goes through brokers (see figure 1). Middle-agents may be considered as an augmentation of the client-server model (see e.g. [14]). Information brokers, however, are capable of performing 2-way tasks: users and sources can act both as clients and as servers. Thus, in ID middle-agents are an extension of a double client-server model.

These 2-way tasks in ID result in a special form of duality, having a *cumulative* nature. It is cumulative over the directions of communication: either from user, via broker to source, or the other way around. Consider, for example, user privacy. Knowledge about user preferences initially is only available in the user himself and can only reach a source through a broker. The dual criterion, concerning source openness, is cumulatively communicated the other way around. Other instances of cumulative duality are considered in section 3.

2.3 Cumulative Duality Matrix

Most specific ID criteria embody a cumulative user-source duality. We propose the *Cumulative Duality* matrix (CD matrix) as an instrument for representing and analysing the influences of cumulative dual criteria upon the design of information brokers. These criteria should be incorporated in a method for agent design. In, for instance, the AWIC method ([10]) the criteria influence the Interoperability and Coordination models.

The general form and characteristics of CD matrices are depicted in Figure 2. The cumulative user criterion is covered by the leftmost column. The dual cumulative source criterion is covered by the top line. The nine cells systematically enumerate the possible combinations concerning the criteria.

The upper left-most cell in Figure 2 represents the most critical situation, since the information broker has no means to deal with neither user nor source criteria. As we go right from here, broker and user agents become capable of dealing with the source criterion. Therefore, a user oriented

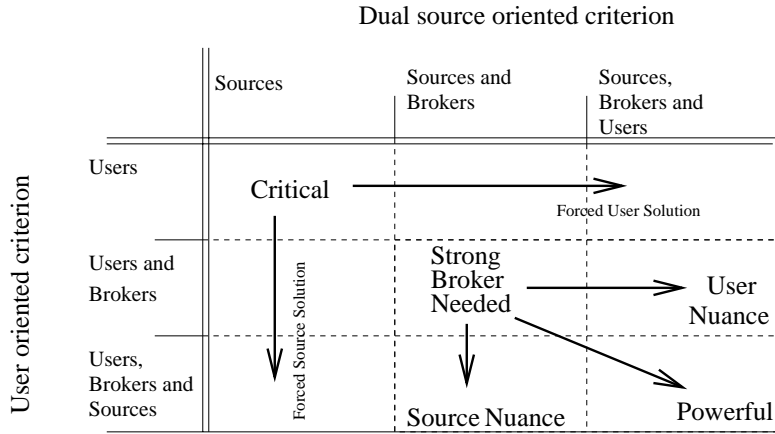


Figure 2: General Cumulative Duality Matrix

solution is forced, i.e. one that demands most of users. If we go down from the first cell, the same occurs with respect to the user criterion.

The middle cell is the minimal cell for which the broker can cope with both the user and source criteria. Since no extra capabilities are available in either user or source, this cell demands a strong broker.

In the cell right from the middle cell, the user also is aware of the source criterion. This means that a nuance, relieving the broker’s burden, towards the user can be made. Analogous remarks apply to the cell below the middle cell. The last cell represents the potentially most powerful situation since both user and source can aid or direct the broker.

The above is easily instantiated with, for example, the user criterion regarding dynamics of user interests. The first row then represents that only the user is able to deal with user dynamics. If also brokers can deal with user dynamics, we find ourselves in the second row, and if sources are equally capable in the third row.

3 Instantiations of CD Matrices

3.1 Partial environmental knowledge

Environmental knowledge in ID concerns the locations, names, or addresses of the agents. Information brokers need this information about both users and sources. If users know the names of sources, they can send addressed

queries (directed IR). If sources know the names of users they can directly send information (directed IF).

Figure 3 provides a CD matrix for the dual location criteria. First we consider the critical situation given in the upper left-most cell. In this situation, the broker initially does not know how to access either users or sources. This means that the broker is passive, i.e. both users and sources have to take the initiative. Processing of queries and incoming documents can only be based on coincidental co-occurrence.

<i>Location of User known by</i>	<i>Location of Source known by</i>		
	S	S+B	S+B+U
U	Passive Br U/S-initiative Coincidence Br	U-initiative (IR) S-querying	U-initiative (IR) Dir S-querying
U+B	S-initiative (IF) U-answering	Proactive Br U/S-privacy	Proactive Br Directed IR Undirected IF
U+B+S	S-initiative (IF) Dir U-answering	Proactive Br Directed IF Undirected IR	Proactive Br Directed IR Directed IF

Figure 3: CD Matrix for Partial Environmental Knowledge

User and source initiative. Next we consider the top row of the matrix in figure 3. In the upper middle cell, the broker can contact sources but no users. This means that users have to take the initiative, leading to a typical IR situation with source querying. In the upper right-most cell, users can perform directed source querying, but still have to take the initiative.

Analogous situations arise in the first column. Here we have sources taking initiative, leading to a typical IF case with user answering. In the bottom cell of the first column, sources have directed user answering on the initiative of the sources themselves.

Proactive brokers. In the central cell of figure 3, the broker is the pivot of the system. Since it knows both user and sources, it can perform proactive

IR and IF, i.e. contacting users and sources on its own initiative. Privacy of location of both users and sources is guaranteed, which also implies that users and sources cannot assist the broker.

In the cell right to the central one, the location of sources is also known by users, which therefore can also perform directed IR. Still, users privacy remains untouched (undirected IF). Analogous remarks apply to the cell below the central one. The location of users is then known by sources but not vice versa, leading to directed IF but undirected IR. The final cell is the only one totally without location privacy, resulting in fully transparent information brokers.

3.2 Privacy of interest and content

When searching the Internet, we would usually like to keep our profile of interest hidden for the outside world. This is an important point in designing information brokers to be treated with care. As an example, in some cases it may be required to keep our search interest private, while in other cases it may help us to make (part of) our interest known to external agents.

In this section we discuss this in the form of two dual criteria. These criteria concern privacy of user interests and source content. The resulting CD Matrix is given in figure 4.

<i>User interests known by</i>	<i>Source contents known by</i>		
	S	S+B	S+B+U
U	Naive Br Broadcaster Assembler	S-overview Hyperindexing U-guide	Selective IR Personalisation
U+B	U-overview Filter topology S-guide	True Broker Total privacy	S-recommender U-privacy
U+B+S	Selective IF Narrowed topol.	U-recommender S-privacy	Arbitrator No privacy

Figure 4: CD Matrix for privacy of interest and Content

First we consider the upper left-most cell of figure 4. In this most critical case, the information broker has no instances of interest and no documents

to match. The only thing left for the broker is to broadcast requests and proposals, and assemble the answers.

User and source overviews. In the middle cell of the top row, brokers can support an overview of the available information. They can, for instance, support Query by Navigation in a hyperindex (see e.g. [5], [4]). In the right cell of the top row, it is possible to perform selective IR, for example by personalising the process of hyperindexing (see e.g. [3]). Similar aspects are found in the left column.

Broker versus arbitrator. The middle cell represents a true information broker, i.e. the broker is the only one knowing about both interests and available content. Privacy for both users and sources is guaranteed. The cell to the bottom represents the broker as introducer or bodyguard. Since the user is unaware of the sources' content, it must be introduced by the broker. In the most powerful combination, the information broker can merely act as an arbitrator since users and brokers are mutually omniscient. This example is taken from [8] and adapted to ID.

3.3 Dynamical Interest and Content

Both user interests and available information are dynamic in ID. User interests are communicated cumulatively towards sources. In the dual case, the content of sources is communicated cumulatively towards users. Information brokers form intermediaries between dynamic parties. See figure 5 for the corresponding CD matrix, where the dual criteria concern the capability of agents to deal with user and source dynamics.

Delayed Interest and Content. In the most critical case, the broker is a static intermediary. This, of course, is no valid solution. Towards a forced source solution, more parties find themselves capable of dealing with user dynamics but not with source dynamics. Eventually, a source stimulated filter is obtained. Here, the source has to provide snapshots satisfying changing user needs since both users and brokers are incapable of doing so.

Interface vs. Assistant. In the middle cell, the information broker is fully responsible for the dynamics of both users and sources. It forms a

<i>Can deal with user dynamics</i>	<i>Can deal with source dynamics</i>		
	S	S+B	S+B+U
U	Static Intermediary	Delayed User Interest Filter ("Static User")	User Monitored Collection
U+B	Delayed Doc Provider ("Static Source")	Dynamic Interface	Dynamic Guide for Sources
U+B+S	Source Stimulated Filter	Dynamic Guide for Users	Dynamic Mutual Assistant

Figure 5: Cumulative Duality Matrix for Dynamical Interest and Content

dynamic interface between users and sources. The cell to the bottom of it represents a dynamic user guide, providing dynamic user access to sources. In the most powerful combination, the information broker can act as a mere assistant providing dynamic mutual support.

4 Other Cumulative Dual Criteria

Dual *time constraints* play an important role in ID. First, in IR, users have short term interests and need direct results, whereas in IF, users have long term interests and do not require results to be delivered immediately. Second, the timeliness of documents is treated differently as well. Where in IR the complete set of documents, including old ones, is queried, IF requires relatively new documents to be sent to users.

Traditionally, a user query or profile only consists of a description of the topic of interest. According to Cooper ([7]) not only topicality (logical relevance) but also *situational factors* (usability) should be taken into account. Dual criteria are obtained with respect to user interests formulated in terms of situational factors as well, and document characterisations supporting these too. The inclusion of situational factors opens the possibility for more interaction in the form of negotiation ([16]).

Agents in the ID paradigm form an *open environment*: agents can come

and go when they want. This requires highly dynamic cooperation strategies without strict commitment ([13]). The electronic (ID) market stresses the importance of supporting open environment negotiation by letting only the most well adapted brokers survive.

Even more cumulative dual criteria can be formulated within ID but also more in general. The consequences of those criteria can be clearly depicted and analysed through the user of CD matrices since these give a structured overview on a conceptual level. This will make the process of broker design better structured and motivated.

5 Conclusions & Further Research

We introduced Cumulative Duality matrices as a mechanism to analyse the *role* of the information broker within the ID paradigm, and of middle-agents in dual paradigms in general. We argued that the duality in goals of users and brokers implied the need for neutral intermediaries.

Design criteria for information brokers were shown to exhibit a cumulative dual nature in Information Discovery. The sections above showed the influence of a number of dual ID criteria on the conceptual design of information brokers. A number of general conclusions can be made.

First, it appears that the information broker must be able to deal with both user and source criteria to obtain a viable broker design. This means that there are only four possible cells, i.e. those in the bottom right corner.

Second, different points of view can be adopted, resulting in different favourable broker designs and focusing on different regions of the CD matrices. The function of the information broker is to provide a transparent connection between different capabilities, needs, and knowledge. In that way the broker has to ensure a basis of equality between users and sources. Two major points of view are (1) user oriented, focusing on the cell right to the middle, and (2) source oriented, focusing on the cell below the middle.

Third, the potentially most powerful cell is in certain respects the most problematic as well. For instance, user privacy is not viable.

Finally, the dual nature of the ID criteria is visible in CD matrices as well. Taking the diagonal from top left to bottom right as a axis, the left bottom corner and right upper corner are mirror images with respect to the dual criteria used.

Further research can be focussed on a number of aspects of CD matrices. For instance on the combination of several CD matrices. The combination of

CD matrices can be researched in the light of more traditional matrices. For example, by the analogy of multiplying and adding (mathematical) matrices. This will lead to the possibility to investigate and analyse hybrid brokers.

In addition, CD matrices can be combined with agent design tools. For instance, they could be included in the AWIC method as described by Mueller in [10].

Finally, the use of CD matrices for personalising brokers can be investigated. For this, we envisage CD matrices as part of advanced user interfaces in which the user can select the most appropriate broker variant.

References

- [1] B. Aoun. Agent Technology in Electronic Commerce and Information Retrieval on the Internet. In *"Proceedings of AusWeb96 Second Australian World Wide Web Conference*, Australia, 1996.
- [2] N.J. Belkin and W.B. Croft. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):29–38, December 1992.
- [3] F.C. Berger and P. van Bommel. Personalized Search Support for Networked Document Retrieval Using Link Inference. In R.R. Wagner and H. Thoma, editors, *Proceedings of the 7th International Conference DEXA'96 on Data Base and Expert System Applications*, volume 1134 of *Lecture Notes in Computer Science*, pages 802–811, Zurich, Switzerland, September 1996. Springer-Verlag.
- [4] P.D. Bruza. *Stratified Information Disclosure: A Synthesis between Information Retrieval and Hypermedia*. PhD thesis, University of Nijmegen, Nijmegen, The Netherlands, 1993.
- [5] P.D. Bruza and Th.P. van der Weide. Stratified Hypermedia Structures for Information Disclosure. *The Computer Journal*, 35(3):208–220, 1992.
- [6] A. Chavez and P. Maes. Kasbah: An Agent Marketplace for Buying and Selling Goods. In *Proceedings of PAAM 96*, pages 75–90, 1996.
- [7] W.S. Cooper. A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval*, 7:19–37, 1971.

- [8] K. Decker and K. Sycara. Middle-Agents for the Internet. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, Nagoya, Japan, August 1997.
- [9] D. Kuokka and L. Harrada. Matchmaking for Information Agents. In Huhns M.N. and M. Singh, editors, *Readings in Agents*, San Francisco, California, 1997. Morgan Kaufmann Publishers.
- [10] H.J. Mueller. Towards agent system engineering. In *Data and Knowledge Engineering*, volume 23, pages 217–245, 1997.
- [11] M.V. Nagendra Prasad, V.R. Lesser, and S.E. Lander. Retrieval and Reasoning in Distributed Case Bases. In *Journal of Visual Communication and Image Representation, Special Issue on Digital Libraries*, volume 7 (1), pages 74–87, 1996.
- [12] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, London, United Kingdom, 2nd edition, 1979.
- [13] T.W. Sandholm and V.R. Lesser. Automated contracting among self-interested bounded rational agents. Technical report, Computing Science Department, University of Massachusetts, Amherst, 1995.
- [14] G. Wiederhold. Mediators in the Architecture of Future Information Systems”. In *IEEE Computer*, pages 38–49, March 1992.
- [15] B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and T. van der Weide. An Electronic Commerce Paradigm for Information Discovery. In P.M.E. De Bra, editor, *Proceedings of the Interdisciplinaire Conferentie Informatiewetenschap*, pages 56–60, Eindhoven, The Netherlands, November 1997.
- [16] B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and T. van der Weide. Opportunities for Electronic Commerce in Information Discovery. Technical report, University of Nijmegen, Nijmegen, The Netherlands, November 1997. Technical Report: CSI-R9722.
- [17] B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and Th. van der Weide. Towards an Agent-Based Retrieval Engine. In J. Furner and D.J. Harper, editors, *Proceedings of the 19th BCS-IRSG Colloquium on IR research*, pages 126–144, Aberdeen, Scotland, April 1997.

- [18] B.C.M. Wondergem, P. van Bommel, T.W.C. Huibers, and Th.P. van der Weide. Agents in Cyberspace – Towards a Framework for Multi-Agent Systems in Information Discovery. In *Proceedings of the 20th BCS Colloquium on Information Retrieval, IRSG98*, Grenoble, France, 1998.
- [19] M. Wooldridge and N.R. Jennings. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.