IST-2001-34825

WP4 – Knowledge-based Retrieval
Task T4.4
Deliverable D4.4

Specification of the Interface Design

(Final, **29/10/2003**)

Abstract – A large amount of strategically relevant business information is contained in unstructured texts. While information monitoring and brokering approaches are used to contextualize such documents and to enrich them by metadata, text mining is used to explore large document spaces. So far, little attention has been paid on a value-adding combination of these technologies. In this deliverable we show how metadata and documents can be complementarily represented and used interactively to support users in text corpus analysis. We specify the UI design and functionality of SEWASIE's integrated monitoring agent and visualisation component, i.e. a text analysis portal which displays inter-document similarity by means of so-called document maps, complemented by a display of the domain ontology and metadata-based access methods. We also show how this component is linked to the user interface of a multidimensional business planning environment based on Online Analytic Processing (OLAP) technology.

Andreas Becks
Fraunhofer-Institut für Angewandte Informationstechnik FIT
Ahornstraße 55
D-52056 Aachen
andreas.becks@fit.fraunhofer.de

1. Document information

Document ID code	D4.4		
Keywords	User Interface, OLAP Annotation, Visualisation		
Current classification of this document	**Final**	Official date of reference	**29/10/2003**
Distribution level	**This document may be distributed to the general public for dissemination and information purposes.**		

Editor	Andreas Becks	FIT
Authors	Andreas Becks	FIT
	Christian Seeling	FIT
	André Brunner	FIT
Reviewer	Giuseppe Santucci	UNIROMA

Version history		
Date	*Version*	*Description*
28/9/2003	Draft 1	First draft of document
7/10/2003	Beta 1	Ready for review
29/10/2003	Final	Revised version according to comments by the reviewer

1.1. Copyright notices

2. Executive summary

In this deliverable we specify the user interface design and functionalities of SEWASIE's integrated monitoring agent and visualisation component, i.e., a text analysis portal which displays inter-document similarity by means of so-called document maps, complemented by a display of the domain ontology and metadata-based access methods. We also specify how this component is linked to the user interface of a multidimensional business planning environment based on OLAP technology. All specification information is given in section 5 by describing layout and functionality of the user interfaces and by presenting the screen design (mocked-up screen shots). Section 6 discusses rationale and philosophy of the visualisation component's UI, points out novel features and paradigms and presents related work. Section 7 clarifies some basic interactive features of the interfaces by presenting a mock-up based on the SEWASIE performance planning scenario.

3. Index of contents

4. Introduction

SME's[1] need to continuously monitor information about customers, competitors, products or market-relevant events in order to assess their situation in a global setting. A great deal of the strategically relevant information is encoded in natural language [18]. Management reports, surveys, and news tickers are only some examples. For the adequate evaluation of company performance and for planning tasks in the light of global markets this kind of external information is crucial and needs to be seamlessly linked to structured planning information often found in business warehouses in order to close the intellectual gap between internal key data and 'soft' external information (see also [24]).

SEWASIE's monitoring agent provides a set of functionalities aimed at supporting these complex tasks, including means for monitoring relevant data sources that deliver text-based background information [2] and a novel functionality for linking semi-structured information (stemming from heterogeneous observed sources) to multi-dimensional business data contained in OLAP[2] cubes using an ontology-based relevance measure [3]. However, the combination of multidimensional data warehouse and text-model-based information monitoring tools within a decision support console – as envisioned by SEWASIE's WP4 – does not only cover the integration of structured and unstructured information at the systems' backend. From an HCI perspective the research challenge is to develop an integrated view on the user interface level that connects internal OLAP data with external text documents, allows to keep track of the information context (*traceability*), and provides means to *explore* the space of collected external information. With the advent of the semantic web the challenge of dealing with structured and unstructured information in integrated visual interfaces has gained much interest (see e.g. [9]). However, up to now only few results exist. In section 6 we present some systems that tackle the problem in some specific settings.

In this deliverable we introduce an UI paradigm that offers a solution to the problem of linking structured planning data and unstructured text information. The user interface of SEWASIE's monitoring agent and visualisation component combines metadata-based information monitoring and text mining in order to help users to exploit both types of information. We also show how this component is linked to the user interface of a multidimensional business planning environment based on OLAP technology. The visualisation interface can also be used to visualise text documents returned in response to a query posed to the SEWASIE's query interface. More specifically, we present a text analysis portal where metadata descriptions of documents (called 'contexts' according to [14]) and inter-document similarity are visualised, enabling the analyst to simultaneously examine documents on a conceptual and natural-language level. Both, navigation and metadata representation are based on a text domain model. This domain model is also linked with hierarchical dimension model of the OLAP world.

4.1. Domain-Models and Multidimensional Data Models: Strangers or Brothers?

For realising the vision sketched above, tools and interface paradigms from two very different areas need to be integrated. This section identifies the similarities between the underlying data models and shows that they can indeed serve as bounding elements in an integrated interface.

The client side of data warehouses, specialised for dealing with internal quantitative business data, is usually based on multidimensional data models as captured e.g. in the CWM[3] standard [27]. In contrast, tools for information brokering and monitoring [12] focus on domain specific and personalised information supply and use domain models that are optimised for describing the context of text documents by means of meta data. However, while research has been carried out separately in the two areas of data warehousing and information brokering/retrieval, both fields surprisingly came up with rather similar modelling structures. In AI such meta models would be called formal semantic networks (or ontologies), the data base world would speak of extended entity-relationship models. Objects, attributes and relationships (in particular specialisation/generalisation hierarchies), and synonym sets play a central role.

Still, in both areas these general, rather complex network structures turned out to be of limited usefulness for end users. Instead, many applications use hierarchical model structures which most user's are familiar with (e.g. from file browsers or folder organisation in document management systems), thus avoiding the well-known phenomenon of feeling lost in hyperlinked network structures. More specific, data warehouse clients

[1] Small and Medium-sized Enterprises

[2] Online Analytic Processing

[3] Common Warehouse Model

typically come along with limited multidimensional data models, consisting of clearly defined base dimensions that reflect the user's interest, and hierarchies (*is-a* or *group-by*) which support different degrees of granularity for data analysis [5]. Similarly, research on human-computer interaction (HCI) has defined so-called overlay models (cf. [13]) which are used to describe user interests, competencies and other situation parameters on top of underlying domain ontologies in a strictly hierarchical manner, complying with typical end user preferences. Following this principle, domain modelling formalisms have been introduced (e.g. [14]) which use category hierarchies as feature dimensions for contextualising documents in information brokering and monitoring environments.

4.2. End User Types and Requirements

There are different classes of typical end users of the monitoring agent and visualisation component. For each user class there is a set of specific requirements for the UI:

User type	Specific requirements
Simple searcher ([23], section 6.4): User entering from the search interface of the SEWASIE portal. This kind of occasional user is only interested in a visualisation of query results in an easily enjoyable format.	Users of this class need an intuitive, easy-to-use interface which visualises the search result and offers some functionality to explore the result set. The user should not be confronted with monitoring-specific interface features or complex analysis tools. A corresponding design specification is given in section 5.3.1, Visualisation of Query Results.
Frequent searcher ([23], section 7.1.2): User interested in long-term monitoring of information. This kind of user has subscribed to a specific domain model and has specified his monitoring profile on top of this domain model.	When exploring monitored information, users need to understand the relationship of that information to their domain and interest model. The context of each information within the models should be accessible at each time. A corresponding design specification is given in section 5.3.2, Visualisation of Monitoring Results.
Business analyst, user of extended functionality ([23], section 7.1.2): OLAP analyst who has also subscribed to long-term information monitoring. This kind of user is a specifically trained analyst who is used to work with multidimensional business information tools.	Same as above. Moreover, the domain model for monitoring should be related to the multidimensional OLAP model in order to ease the switch between OLAP and monitoring tool. A corresponding design specification is given in section 5.3.3, Visualisation of Annotation Results.

Concerning the second and third type of user, a guiding principle for the interface design should be to juxtapose the explicitly modelled, structured metadata on the one hand and weakly-structured text information on the other hand. Users should be able to find answers to questions like 'How are these classes related with respect to the text collection?' or 'How are these semantically similar documents characterized in the model of the domain?'. Therefore, the handling of structured and unstructured information should be equally weighted concerning the richness of functionality and the fraction of display in the user interface. Views on both types of information should be tightly integrated so that each navigational step in one view is simultaneously presented in the complementary view. To prevent the analyst from being overwhelmed with information, he should always be given an overview of both, the metadata model (domain ontology) and the document space. Concerning functional requirements, it is widely accepted to support the core analysis activities of Shneidermans' visual information seeking mantra [21]: overview, filter, and details on demand.

4.3. Elements of an Integrated Decision Support Console

Following the requirements from the last section, mandatory elements that make out a comprehensive analysis support are worked out now, covering design options as well as decisions made for our prototype.

Firstly, capabilities for the representation and exploration of the integrated hierarchical data model are needed. There are many scientific and commercial options for the choice of information visualisation metaphors for metadata or documents. Comprehensive discussions of potentials, drawbacks, metaphors and task-adequacy can be found in [7], [19]. In the area of visualisation of textual metadata, there is emerging interest in metaphors which allow the user to browse through taxonomies [4], [16]. Accordingly,

commercial tools become more and more advanced. Due to fact that most users are fairly familiar with linear representations of hierarchies, we chose to deploy such an explorer-like metaphor. These representations are known to be concise and intuitive for the navigation through hierarchical models (cf. for example [9]).

Secondly, a means for representation and analysis of semi-structured document collections, comprising the combined examination of metadata and natural-language text content is needed. Concerning the visual assistance for the exploration of text documents, text clustering is an important topic, and a variety of clustering metaphors are available (e.g. [4], [17], [15]). In our prototype we use a Kohonen-based clustering map approach as a metaphor for the visualisation of documents [4]. Once calculated, the display shows the inter-document and inter-cluster similarities at the same time (similarity-values can be computed from documents or metadata, alternatively). The component used has proven its usefulness in various case-studies and user experiments. For the metadata associated with documents, value-tables are used.

Thirdly, the interactive combination of the workspaces has to be accurate and well-balanced. Recent approaches address the combination of text analysis and model-based navigation (cf. for example [9]) but are obviously putting emphasis on one of the topics and subordinating the other. This, however, causes a lack of interaction functionality. Aiming to reach a balance (with respect to functional capability and portion of the user interface) between navigation through the integrated data model and exploration of the document annotation space, we decided to present and update both workspaces in parallel, i.e., the interaction in one workspace is mirrored in terms of the respective other metaphor, complementing the selections.

5. Specification of User Interfaces

In this section we present the **functionalities** and the **layout** of the different user interfaces of the monitoring agent and visualisation component. A mock-up clarifying the main functionalities is presented in section 5.3.3. The interfaces presented here are part of the SEWASIE portal which comprises several modules and module transitions (for details cf.[8], section 10). More precisely, this portal will be web-based and organised in tabs corresponding to the main modules. The user can switch between the tabs (stateless transition) or transfer data from one module to the other (state transition). The following table summarises the interfaces generated in WP 4 and specifies their entry points according to the SEWASIE portal concept.

User interface	Component	Entry point	Transition type
Monitoring interface	Monitoring agent	SEWASIE portal tab	stateless
OLAP reporting interface	OLAP tool	SEWASIE portal tab	stateless
Query result visualisation	Visualisation component	Query results interface	state transition; data handed over includes URI list and similarity matrix (if available)
Monitoring results visualisation	Visualisation component	Monitoring interface	state transition; data handed over includes the user's monitoring profile
Annotation results visualisation	Visualisation component	OLAP reporting interface	state transition; data handed over includes URIs of annotated documents and the user's monitoring profile

5.1. User Interface of the Monitoring Agent

This dialogue is activated if the user enters the monitoring interface tab of the SEWASIE portal. Here, the user can access the repository of monitoring results. Figure 1 shows the different filters that can be combined interactively to select the most interesting monitoring results. The domain model is represented as a tree where the user can select categories of interest. The checkbox *Changes Only* allows the restriction to news, i.e. the 'delta' with respect to the monitoring results which have already been shown before. The checkbox *Last Version Only* strives for reducing the results by eliminating all duplicates in favour of the last version of a document. This is particularly important for documents which are changed frequently. Last but not least, the result list can be constrained along the time dimension to show up only actual results.

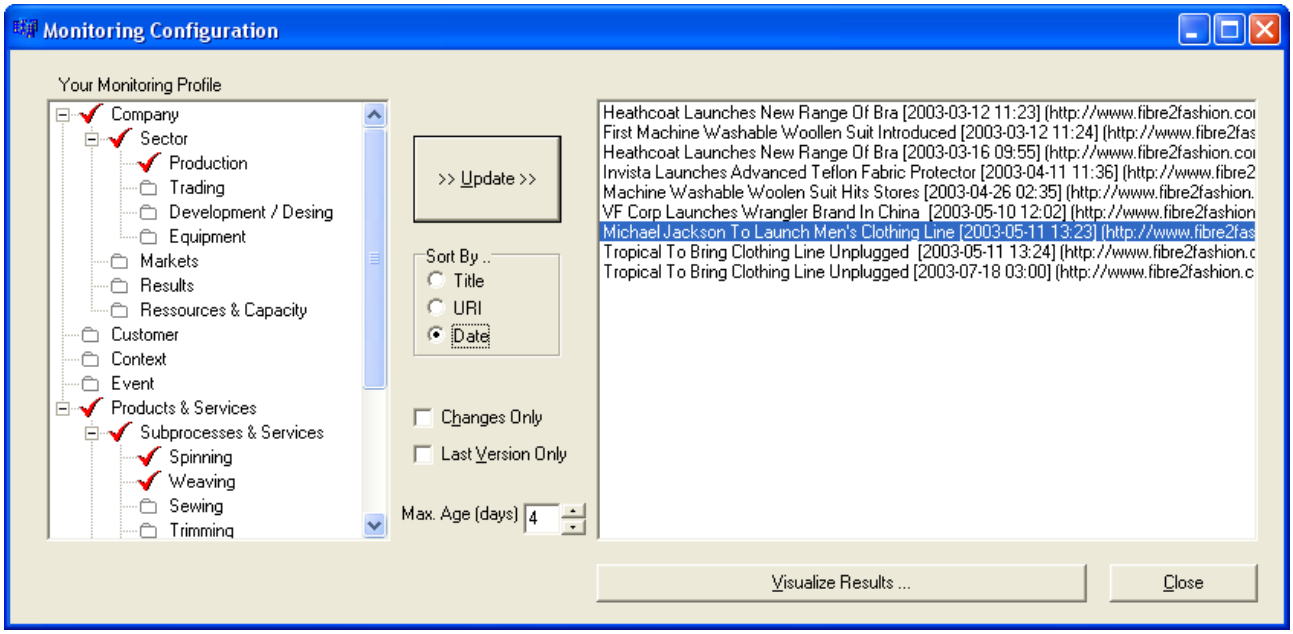


Figure 1: Screenshot of the Monitoring Agent UI.

Monitoring results are displayed in a result list which comprises document title, URI and date. This information can be used as sorting criteria. For deeper investigations of long lists of results the user might decide to visualise the results. Clicking the respective button initiates the calculation of a document map and leads to the interface for visualising monitoring results of the visualisation component (see section 5.3.2).

5.2. OLAP Reporting Interface

This interface is activated if the user enters the OLAP tab of the SEWASIE portal. The Web interface of the OLAP tool basically comprises standard OLAP functionalities (cf. D4.1 [2], Appendix A.3). With respect to the monitoring and visualisation component developed in SEWASIE the following extensions at the user interface level are relevant:

a) **Button for document annotation:** If this button is pressed (cf. Figure 2), a request for annotating the displayed OLAP report with documents from the monitoring agent's document repository is automatically generated and executed in the background. An annotation result screen is displayed that presents the result of the annotation request.

b) **Annotation result screen:** This screen (cf. Figure 3) displays a list of those documents from the monitoring agent's document repository that where assessed as relevant for the OLAP report from which the request was started. The relevance is computed according to the relevance measure presented in [3]. For each document assessed as relevant the list contains an entry consisting of the document's title (if available), the relevance value computed by the relevance measure [3], the original URL of the document as well as the URL of the document in the monitoring agent's cache, and a list of those OLAP dimensions of the report for which the document is most relevant (cf. [3] for details). The user can view documents by clicking on the original or cached URL. If he does so, a separate browser window displaying the document will open. By pressing the OK button the annotation result screen will be closed and control will be handed over again to the report screen from which the request was started. If the user presses the "show in SWAPit" button the UI for visualising annotation results will be started (cf. section 5.3.3).

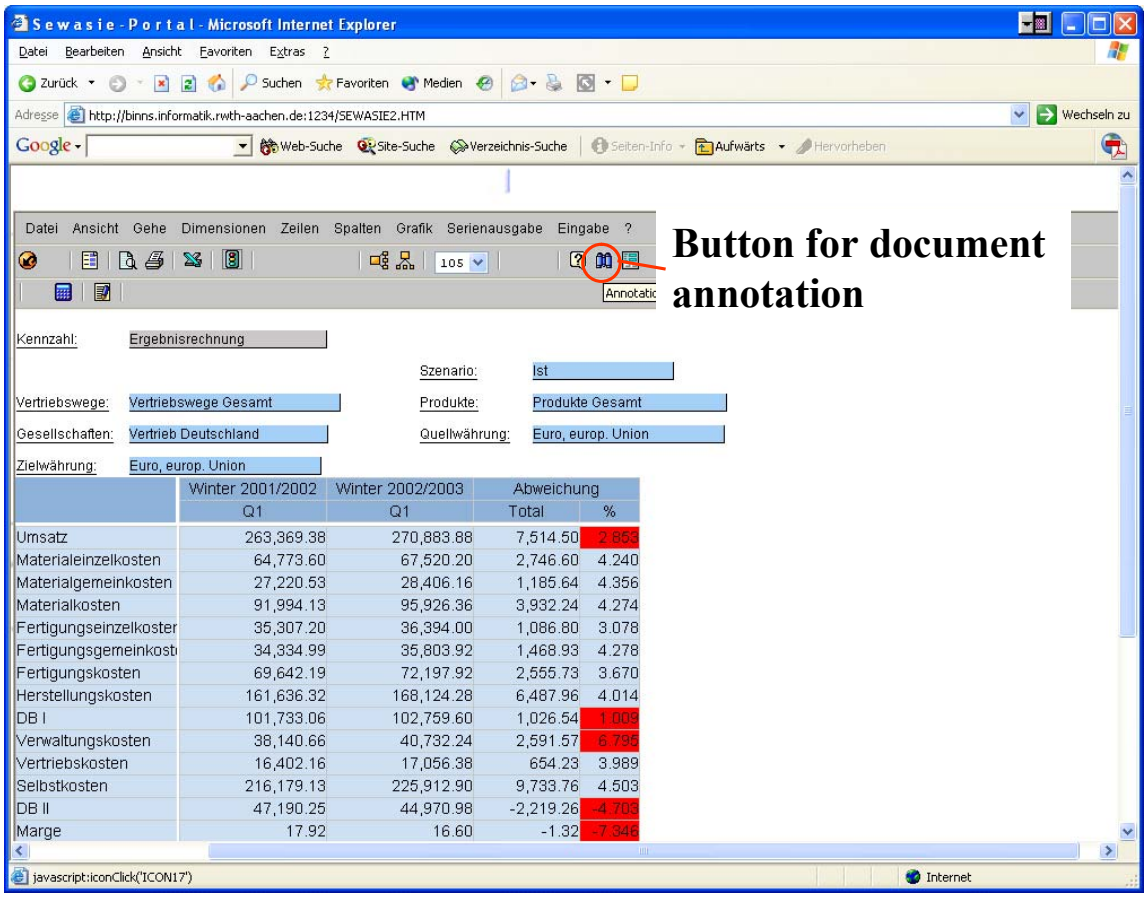


Figure 2: OLAP UI extended by button for document annotation, depicted by binoculars icon

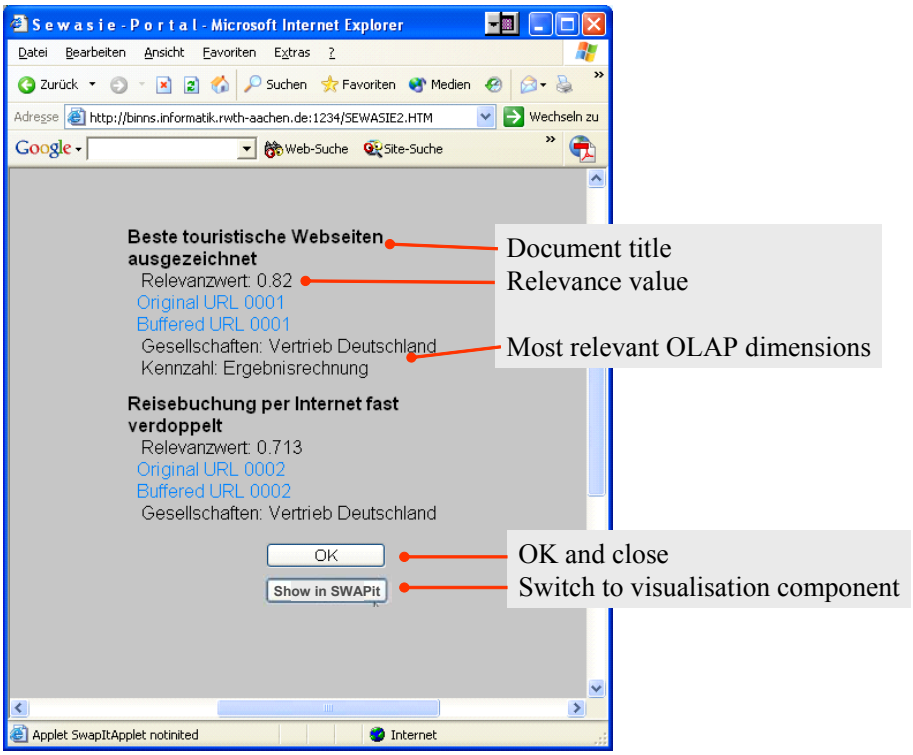


Figure 3: Annotation result screen

5.3. User Interfaces of the Visualisation Component

The user interface of the visualisation component will be referred to as SWAPit[4] in the following. SWAPit is a novel interface paradigm developed in SEWASIE that combines similarity-based navigation in so-called document maps with ontology-based navigation over document categories. This capability will be fully-fledged in the interface mode for visualising monitoring and annotation results (sections 5.3.2 and 5.3.3).

5.3.1. Visualisation of Query Results

This mode of SWAPit is activated by the query interface if the user requests to visualise the query result. There are two modes of displaying documents, depending on the information handed over to the visualisation component:

a) If a matrix specifying the similarity of each pair of documents in the result list is given (e.g. based on the documents' metadata), SWAPit will display the similarity structure of result sets containing multilingual data.

b) If a list of URIs is given without specifying the mutual similarity of items in the list, SWAPit will use its linguistic indexing and similarity computation. In this case the similarity structure of monolingual[5] subsets of result lists will be displayed.

Based on the respective information handed to SEWASIE's visualisation component, SWAPit automatically computes a document map which shows the similarity structure of the given set of documents (cf. Figure 4). More precisely, the semantics of this map is determined by the following metaphor: **Documents are represented as red dots in the map**. Similar[6] documents are grouped as neighboured points, located in common bright shaded areas. These areas are separated by dark borders, representing the distance between document groups. The darker the colour, the more dissimilar are the separated groups of documents.

[4] See section 6 for a discussion of the name.

[5] The current implementation comprises linguistic methods for German and English.

[6] Document similarity is determined by the similarity measure applied for computing the similarity matrix (case 'a' from above). If the built-in method of SWAPit is used the similarity is computed based on the documents' keyword distribution (case 'b').

For the visualisation of query results, SWAPit offers the following basic functionalities in the tool area at the lower left panel:

- As the **basic functions** of the document map workspace, **documents can be opened by point-and-click**. The opened document will be displayed in the integrated Web browser the "**browse" tab**. The user can **select arbitrary subsets of documents** for further processing by drawing one or multiple selection frames with the pointing device[7]. The selection frames are
- The **search tab** allows posing keyword-based requests against the visualised text collection. Supported search modes are Boolean and full-text search. Query results are not only displayed as ranked list but also highlighted in the map display (yellow rectangles surrounding icons of matching documents). Within the search tab the list of documents matching the query is displayed along with a bar chart visualising the relative relevance value for each document in the list (cf. Figure 5).
- The **term statistic tab** shows the distribution of statistically relevant terms of the document group selected by the current selection frame. There are different term relevance measures: (a) 'significant' terms are determined by the sum of normalized term occurrences in the set of selected documents, i.e. they are the most frequent terms in the group; (b) 'characteristic' terms are more significant in the selected group than in other groups; (c) 'group profile' terms are characteristic terms which occur in most of the selected documents. The list of terms with the highest ranking in the selected profile is displayed in the 'term statistics' tab along with a bar chart visualising the relative weight of each term according to the term relevance measure (cf. Figure 5). If the users clicks on a term in the list document icons of documents containing this term are highlighted in the map display.
- The **browse tab** is used to display opened documents in a hypertext browser.
- The **info tab** shows version and copyright information.

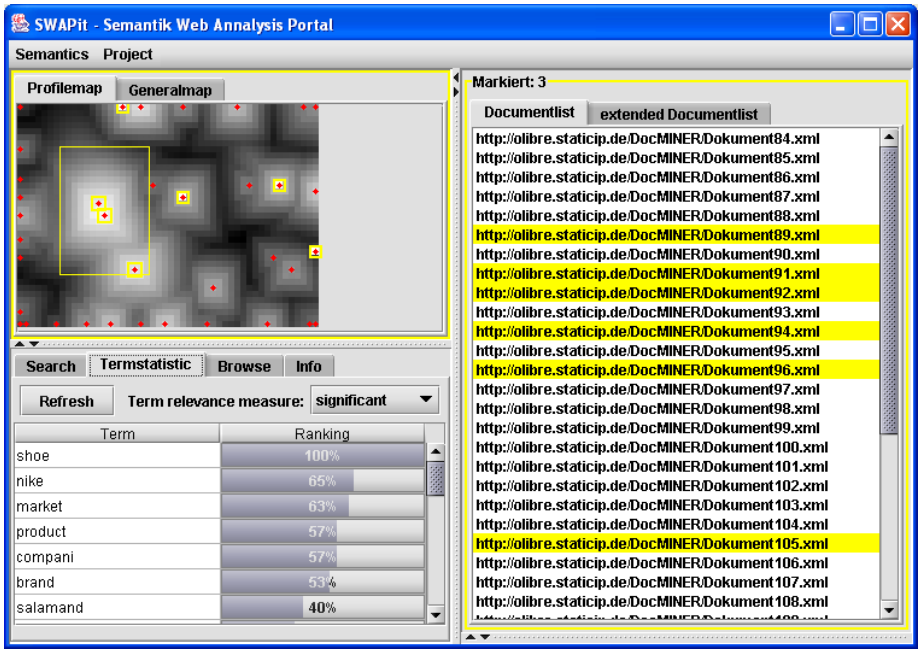


Figure 4: SWAPit mode for visualising query results. *Upper left:* document map with active selection frame (yellow rectangle) and highlighted documents that match a query formulated using the search tab in the tool area; *lower left:* tool area with active term statistics tab; *right:* list of document URIs; that match the query formulated using the search tab are highlighted in yellow.

[7] Multiple selection frames can be drawn by pressing the CTRL key while drawing frames with the pointing device.

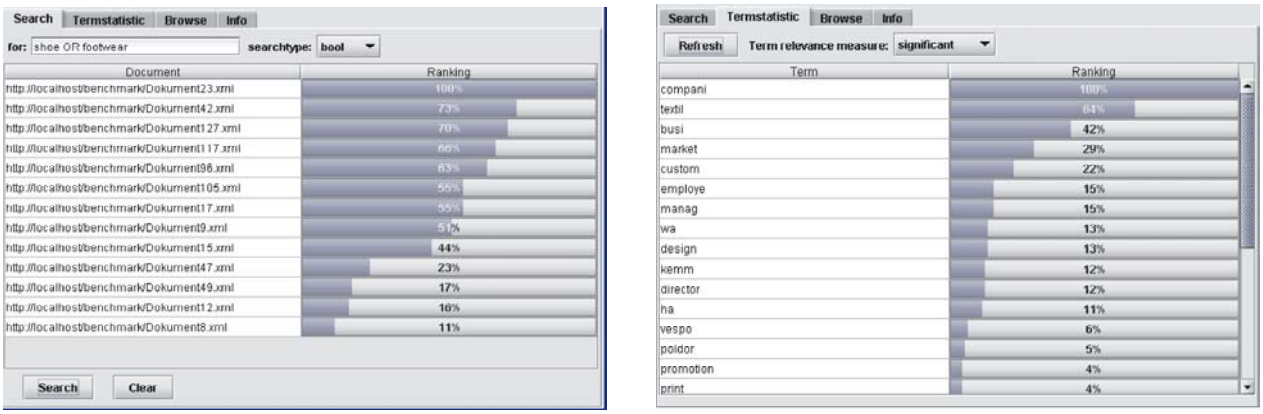


Figure 5: Tool tabs: search tool (left) and term statistics tool (right).

5.3.2. Visualisation of Monitoring Results

This mode of SWAPit is activated if the user enters the monitoring interface via the monitoring interface tab in the SEWASIE portal (cf. section 5.1) and presses (possibly after some configuration activities) the "visualise results" tab. Then, SWAPit will start up in the layout depicted in Figure 6, consisting of the following four components:

- A **document map** displaying the inter-document similarity of documents in the MA's repository that contains the monitored documents. The document map can be used for similarity-based navigation and exploration of documents (details of the map metaphor are given in section 5.3.1).
- A **list of documents** containing URIs and additional metadata (if available).
- A **domain ontology tree** for metadata-based navigation that enables the user to select documents by their (multi-)classification. The domain ontology corresponds to the domain model used by the monitoring agent for monitoring and categorising documents.
- A **tool area** offering a search interface, term statistics, document viewer (hypertext browser) as presented in section 5.3.1. The 'OLAP model' tab is only activated in the 'visualise annotation result' mode presented in section 5.3.3.

Both, domain ontology tree and document map can be used to navigate through the document collection or to select a set of documents which may be analysed using the term statistics tool from the tool area. Navigation means selecting individual documents or groups of documents. This selection is visualised in each of the two panels based on the following principle (see also Figure 6):

- When the user **selects documents in the document map** (by drawing one or multiple selection frames, see section 5.3.1) **or in the document list** (by clicking on one or multiple rows[8] in the list) filters are immediately applied that select and highlight documents in the different panels as follows:
  - The **set of selected documents** contains all documents that correspond to the document icons within the selection frame(s) in the document map, or that have been explicitly marked in the document list, respectively.
  - In the **document map**, **icons** of selected documents are **highlighted by yellow squares** surrounding each document icon.
  - In the **document list**, **rows** corresponding to selected documents are **highlighted in yellow**.
  - In the **domain ontology tree**, **category names** of those categories are **highlighted in yellow** that are assigned to at least one of the selected documents by the monitoring agent (see [2] for details). In addition, the **colour intensity** of the highlights indicates the relative frequency to which highlighted categories have been assigned to the selected documents: The more documents in the selected set are contextualised with a specific category the more intense is the yellow highlighting of that category name.

[8] Multiple selection of rows can be done by pressing the CTRL key while clicking.

- When the user **selects categories in the domain ontology tree** (by clicking on one or more category nodes in the tree view[9]) filters are immediately applied that select and highlight documents in the different panels as follows:
  - The **set of selected documents** contains all documents that are contextualised with at least one of the selected category names by the monitoring agent (see [2] for details).
  - In the **document map**, **icons** of selected documents are **highlighted by red squares** surrounding each document icon. In addition, the **colour intensity** of the highlight indicates the relative number of categories that are used to contextualise the highlighted document: The more of the selected categories are assigned to a specific document the more intense is the red highlighting of that document icon.
  - In the **document list**, **rows** corresponding to selected documents are **highlighted in red.**
  - In the **domain ontology tree**, selected **category names** are highlighted in red.

In order to help users to better recognize the colouring scheme the document map panel and the document list panel are surrounded by a **yellow rectangle** while the panel of domain ontology tree is surrounded by a **red rectangle**. Selected documents can be opened by double clicking on the highlighted document icon in the document map or on the respective document row in the list of documents.

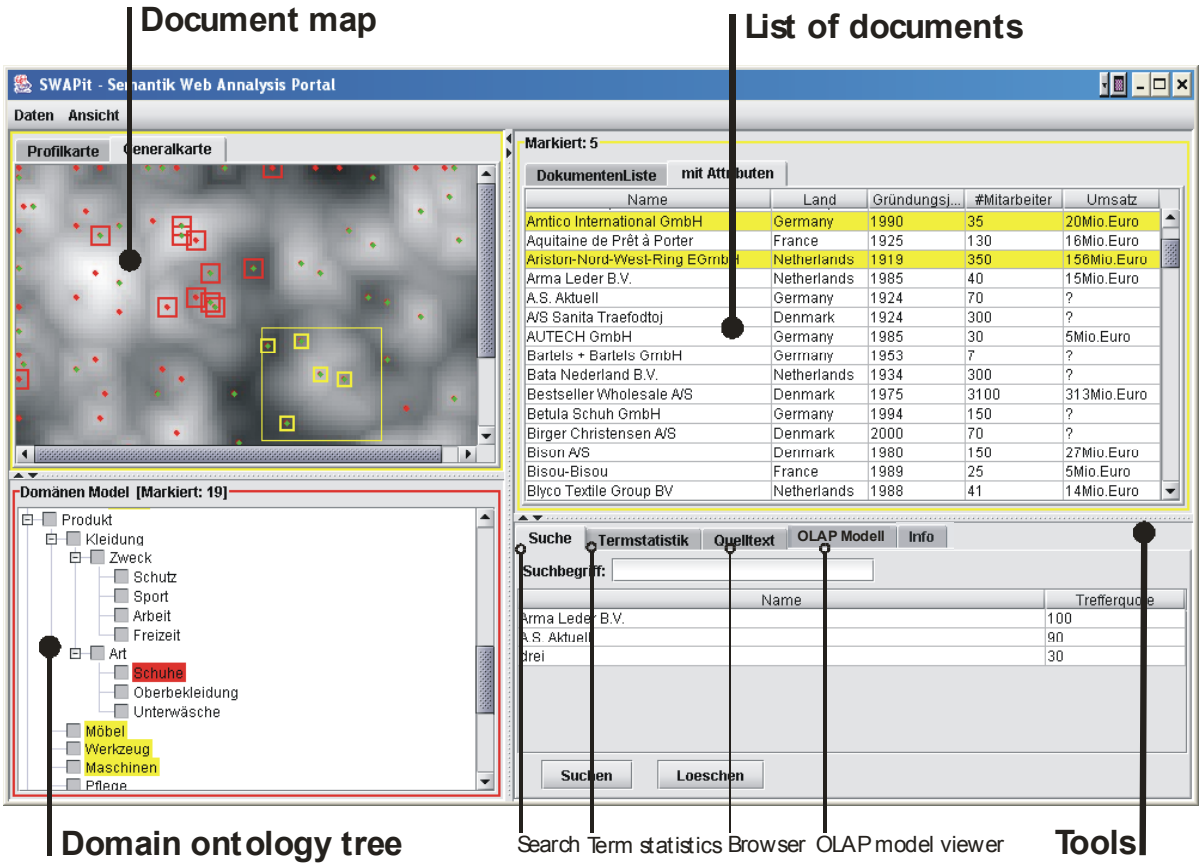


Figure 6: Visualisation of monitoring results

5.3.3. Visualisation of Annotation Results

SWAPit will start with the layout presented in Figure 7 if the user has pressed the “Show in SWAPit” button in the OLAP user interface (cf. section 5.2). In this case:

a) the documents from the list in the annotation result screen of the OLAP interface will be highlighted in SWAPit’s document map area by **flag icons**,

[9] Selection of one category is done by clicking on the category node with the pointing device. A multi-selection can be done by pressing the CTRL key while clicking with the pointing device. If the user selects a category which contains subcategories all subcategories are automatically selected. However, the user can dis-select one or more subcategories explicitly by clicking on the respective category names (while pressing the CTRL key).

b) the "**OLAP model" tab in the tool area will be activated** at start-up, focussing on the dimensions of the OLAP report from which the annotation request was started, and

c) a **button for jumping back to the OLAP tool** will be included. If the user presses this button, control will be handed over again to the OLAP tool which will carry on from the state it was left when the user pressed the button for document annotation (cf. section 5.2).

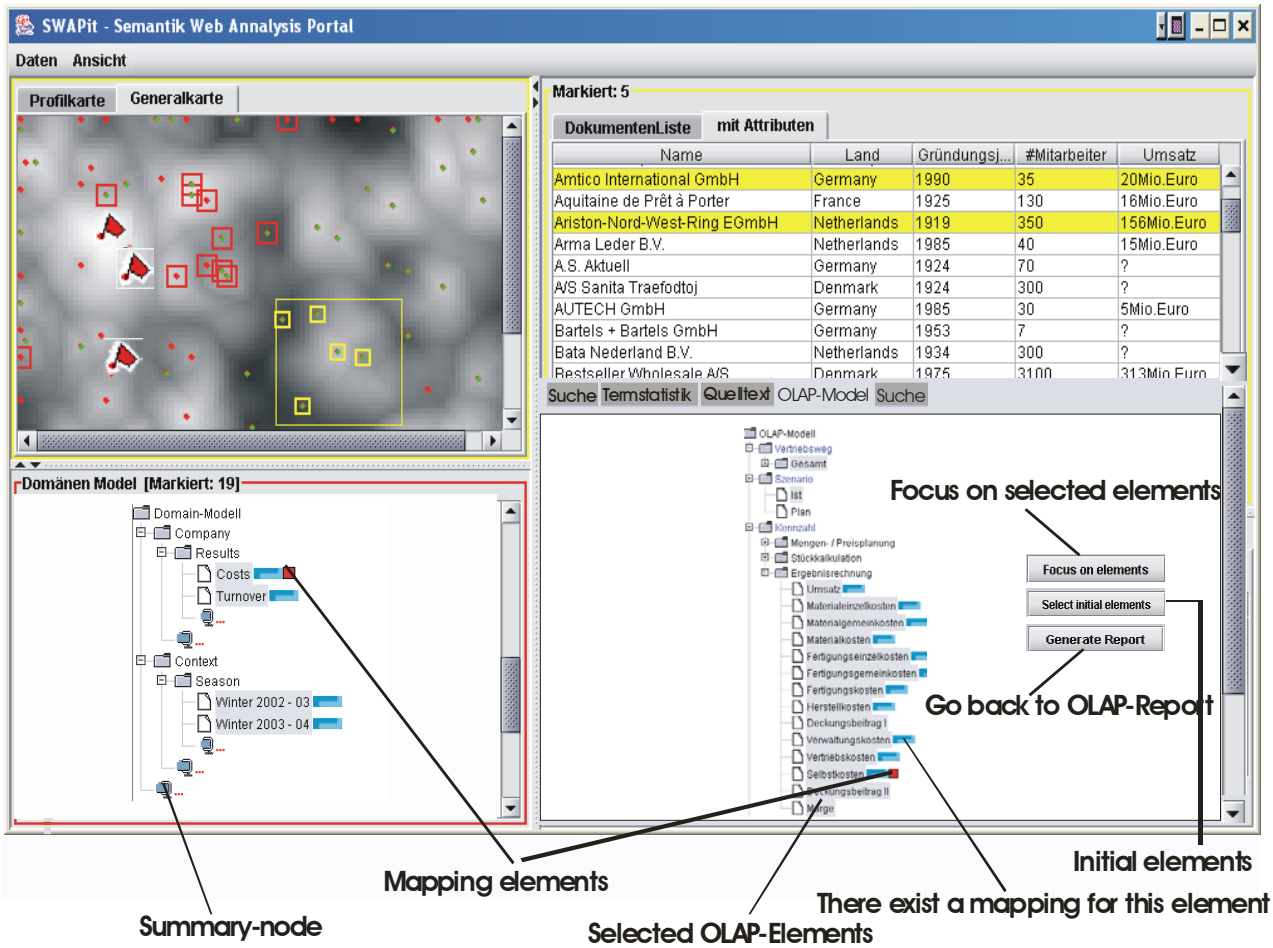


Figure 7: SWAPit showing annotation results. Documents annotated to the report are highlighted by red flags in the map display. Report-relevant elements of the OLAP model are highlighted in the domain ontology tree and in the OLAP model display.

Basically, layout and functionalities are the same as described in section 5.3.2. However, special mark-ups and additional functionality of the domain model view extend the standard behaviour for visualising monitoring results:

- The **"OLAP model" tab** contains a tree that represents the OLAP model of the OLAP tools from which SWAPit was activated. At start-up all the dimensions of the current OLAP report from which the annotation request was started are in focus. With the aid of the mapping between the domain model and the OLAP model all the corresponding elements of the domain model are selected as well[10].
- In the **domain ontology tree**, a **blue icon** behind the elements indicates whether there exists a corresponding element in the other model. To see such a correspondence the user can move the mouse over an element and a **red icon** indicates which element in the other model is affected.
- Starting from the initial selection it is possible to navigate *both*, the domain ontology tree and the OLAP model tree. The behaviour is the same as described in section 5.3.2 for selecting elements in the domain ontology tree *if there is a mapping between both models*. Elements from the OLAP

[10] Actually, this is a subset of all mapping elements.

model that have no mapping element in the domain model cannot be selected for navigation. Whenever the selection changes the new mapping between the models is visualised in both trees.

- *Because the selected elements could be distributed all over the tree, it is not possible to have all selected elements in focus at the same time. Therefore both trees can be optimized by pressing the "**Focus on elements**" **button**, having the effect that all the siblings of a selected node are summarized and represented by a **summary node.** When the user selects such a summary node the represented siblings are expanded again. If a node that is represented by summary node is affected by a new selection (i.e. it maps to the new selection) the summary node is deleted and the node and it's siblings are expanded.*

- At any time the user is able to compare the actual selection with the initial selection. By pressing the **"Select Initial Elements" button** the affected elements are highlighted so that the user can compare them with the other elements.
- With help of the **"Generate report" button** the user is able to go back to the OLAP-tool. The currently selected elements in the OLAP tree are presented in a list and the user has the possibility to generate a new report.[11]

[11] This function is optional and may not be implemented in the SEWASIE prototype.

6. Visualisation Component: UI Philosophy and Related Work

The benefits and novel features of the visualisation component's UI (as presented in section 5.3) shall be discussed briefly in this section. The visualisation component is based on the interactive document map system DocMINER, developed at Fraunhofer-FIT, that has been conceptually extended and technically re-designed for SEWASIE[12]. Document maps visualise the similarity of documents and groups of documents (according to a given similarity measure) by a graphical metaphor reminiscent of geographical cartography [4]. As a text mining method they are used for exploring text databases in many knowledge management applications, e.g. patent analysis, idea management, quality management of technical documentation. The major strength of document maps is that they convey a general picture of a document collection's structure, helping users to grasp the collection's content structure, to relate similar information, to find unique pieces of information, and so forth. Exploration in this sense is mainly done by similarity-based navigation.

6.1. Query Result Visualisation

The user interface for query result visualisation adopts the basic functionality of DocMINER in order to help users to better exploit text information retrieved by SEWASIE's query agent. Similar pieces of information are grouped by the document map, supporting users to understand the structure of the query result set. Thus, the user interface for query result visualisation is a value-adding service on top of SEWASIE's query interface. Displaying the structure of query result sets has a tradition in information retrieval since the mid-90s, aiming at providing visual methods for a better identification of relevant documents in query results (e.g. [1],[10],[11],[20],[22],[25]). Query result sets are usually presented as ranked lists of references to matching documents. Result set visualisation reveals some sort of structure of result sets, making them accessible more intuitively and improving retrieval effectiveness. In particular, a basic assumption is that outliers, i.e. documents wrongly judged relevant, can be identified more easily by visualising the relationships of retrieved items.

6.2. Monitoring and Annotation Result Visualisation

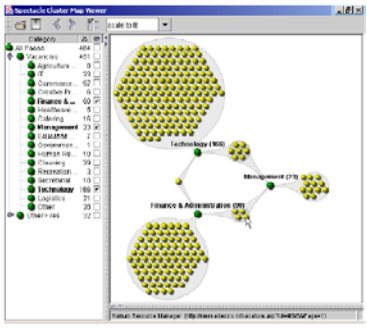
The user interface for monitoring result visualisation significantly extends the document map functionality and introduces novel interaction concepts. The basic idea of the interface design is to *integrate similarity-based and catalogue-like navigation* for text collections. Users can *swap* between both navigation styles at any time. This allows to dynamically contrast and juxtapose document *context* (in terms of metadata) and document *content*. We have chosen the name SWAPit for this new paradigm both as an invitation to the user and partly as an acronym for Semantic Web Analysis Portal. SWAPit supports the three core analysis activities (overview, filter, and details on demand) according to Shneidermans' visual information seeking mantra [21]:

1. For getting an **overview** of the contents of a document group SWAPit provides the term statistics tool which extracts natural language-oriented descriptions from the documents *contents*. These are juxtaposed to the metadata *context* of the same group by concept patterns that are highlighted in the ontology tree and displayed in the attribute-value table of the document list panel. Search results and term distribution are highlighted in the document map, thus presenting their global context. A special case for overview maintenance is realised in the OLAP model tab that presents an hierarchical dimension model often used in business intelligence suites, especially in OLAP applications. This feature preserves the overview and context specifically for users who entry SWAPit from the OLAP tool and will switch back to OLAP analysis after they explored interesting information.
2. The user can **filter** information according to explicit interests by full-text searching, selecting similar documents by navigation in the map display (filtering based on similarity), or interactively configuring a constraint on the domain ontology by selecting classes in the classification tree (filtering based on ontology). Moreover metadata values in the document list panel can be grouped and searched.
3. For receiving **detailed information on demand** users can open and explore documents in the browser. When opening a document the document's context is highlighted in the domain ontology tree. Moreover, the user can view metadata attached to documents in the metadata tab of the document list.

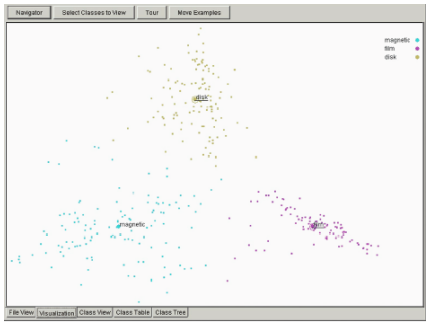
[12] See [2] for information on the technical design.

In summary, SWAPit supports users in making extensive use of the available metadata. Both navigation and analysis are always related to the domain ontology and structured numerical or textual metadata can be analyzed by browsing, characterizing, and searching. The UI treats metadata context and document content equally in both, the extent of functional support as well as the fraction of the display of context and content. The overview is granted since both workspaces (document map and domain ontology tree) are always visible in parallel and updated simultaneously on changes. The focal subset of interest can be inspected in detail or explored in relation to the rest of the collection. The representation of classes and documents has deliberately not been mixed in the same graphical display to avoid confusion. Instead, we designed interactive features that bridge the gap while preserving the context for the user.

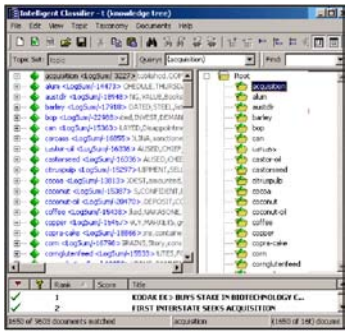
In research up to now only little attention has been paid to interfaces which combine the analysis of structured and unstructured data. Some recent systems visualise both, the structure of a document repository (solely derived from the unstructured text parts) and the metadata, i.e. document attributes and/or taxonomies which can be (i) extracted from the documents, (ii) predefined, or (iii) a combination of both (cf. Figure 8): Spectacle [9] maps document classes and documents to a single graphical display. A predefined hierarchy of non-disjoint classes (taxonomy) is displayed separately and serves to configure the graphical map where groups of equally classified documents are arranged in clusters. Though the intuitively usable tool comes with a smart visualisation it does not give any information about the similarity of documents. Furthermore, with increasing number of documents and classes, it becomes difficult to keep the overview. Though Spectacle allows querying for both, metadata and documents, document analysis is reduced to presenting class membership. Similar to Spectacle, the system eClassifier [6] maps classes and documents to a single workspace. It incorporates multiple algorithms for the generation, analysis, modification, maintenance, and visual representation of taxonomies. The tool offers different views on the derived taxonomies comprising statistical (bar graphs, spreadsheets, plots) and structural representations (trees). Moreover, statistical trend and correlation analysis is supported. eClassifier mainly focuses on the automated extraction and interactive examination of the inherent structural characteristics of document collections. Verity K2 Discovery Tier [26] integrates metadata and document analysis but puts emphasis on advanced retrieval. Navigation and querying are possible for both worlds (the combined search over document metadata and content is called *Parametric Selection*). For a structured access to the document repository K2 offers an iterative procedure: Identify a class, then examine assigned documents and refine the query accordingly.



SPECTACLE [9]



ECLASSIFIER [6]



VERITY K2 [26]

Figure 8: Screenshots of related systems

To sum up, all three tools are putting emphasis on one of two aspects: Analysis of classification and document metadata on the one hand or content-oriented analysis of the structure of the document space. Consequently, the tools are lacking a balanced support and tight interactive coupling of taxonomy and metadata view and visualisation: With Spectacle the exploration of inter-document relations which are not reflected by the taxonomy or metadata are hard to explore. eClassifier does not couple the centroid-based visualisation with the taxonomy; therefore the inter-class relationships are not simultaneously visible. In K2 inter-document relations can be derived on the level of document groups by clustering techniques but are visualised only in list form; the tool could benefit from a graphical representation.

7. Scenario-Based Interface Mock-Up

This section presents a user interface mock-up based on the performance analysis and planning scenario (see also D4.2 [3], section 2), illustrating a selection of UI features of the OLAP tool and visualisation component.

7.1. Entry Point: Performance Analysis in OLAP Tool

A medium-sized German textile retailer analyses the company performance by looking at the statement of earnings in his OLAP system. The OLAP traffic lighting indicates a weak increase of turnover and a strong decrease of margins (Figure 9). This triggers the analyst to search for information on how his data relates to the market. He thus requests to annotate his report with background information from the monitoring repository by pressing the binocular icon in the tool bar of the OLAP UI. As a result the annotation result screen pops up, showing two documents about Hugo Boss. The first text says that Boss intends to keep turnover constant while increasing profit (Figure 10). This attracts the attention of the analyst who now decides to go a little bit more into detail with exploring available background information, therefore pressing the “Show in SWAPit” button.

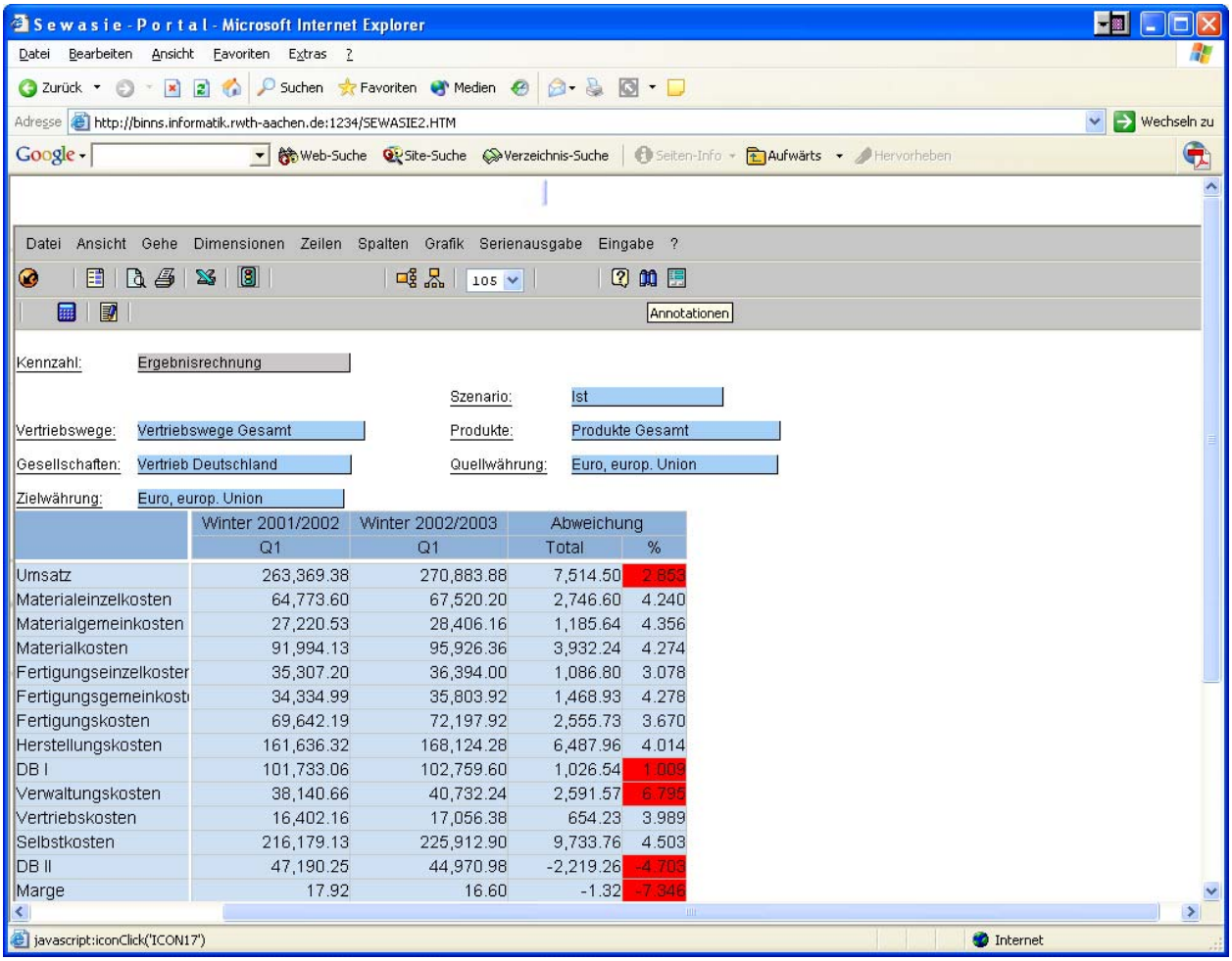


Figure 9: OLAP UI with report

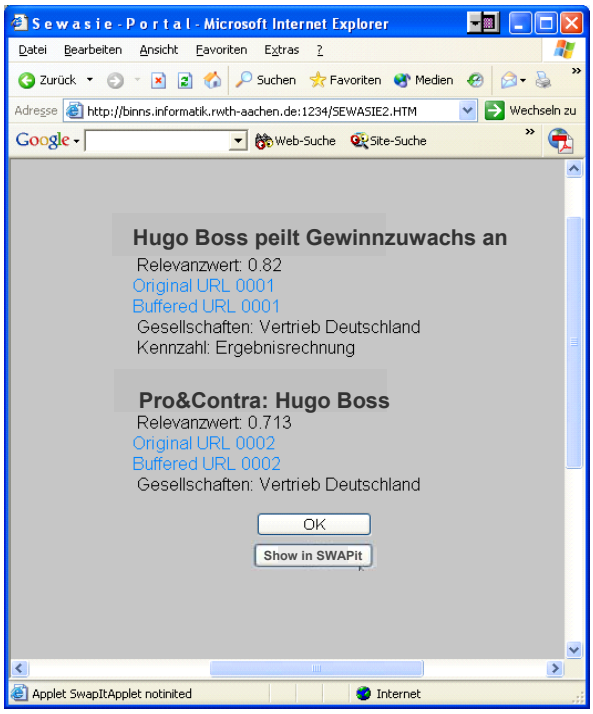


Figure 10: Annotation result list

7.2. *Roundtrip: Exploring Related Information*

Now the SWAPit interface pops up in the SEWASIE portal. The two documents from the annotation result list are marked by red flags in the document map. The user select the two documents by drawing a selection frame in the document map. Immediately the names of the documents in the document list and the domain model categories of the documents in the domain ontology tree are highlighted. Both documents are contextualised with “costs” and “turnover” (Figure 11).

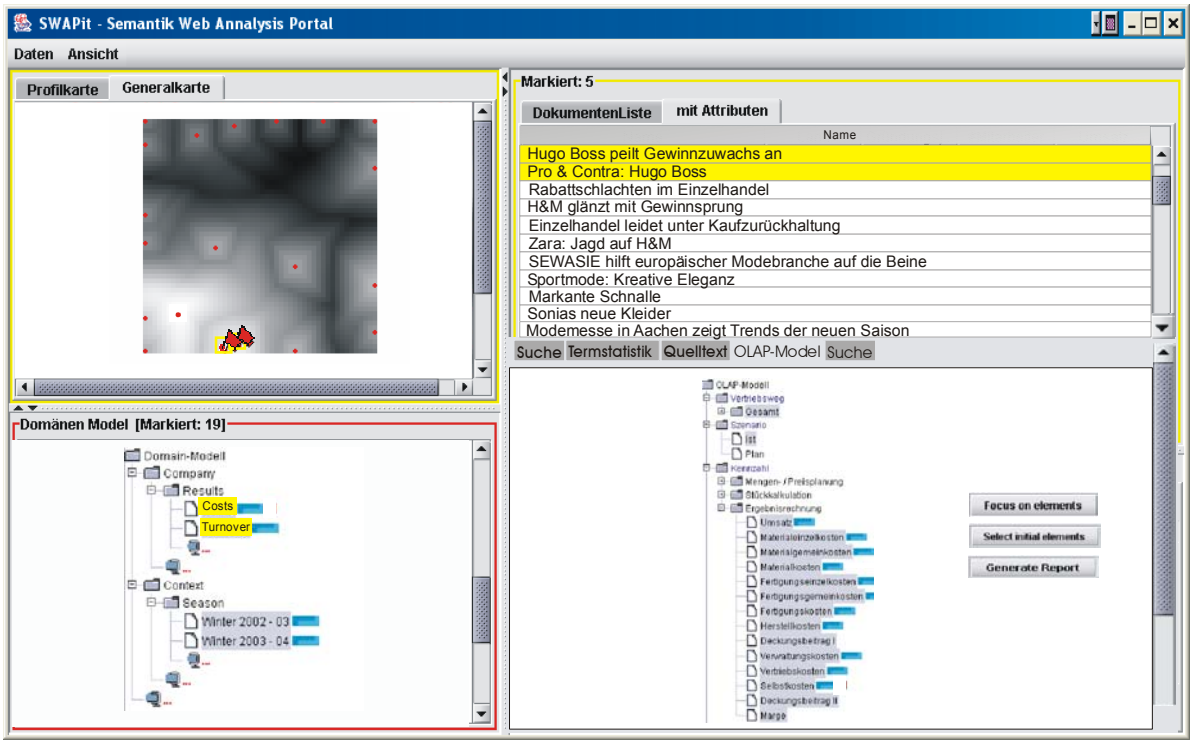


Figure 11: SWAPit interface at start-up time, selection of documents in the map

The user explores similar documents in the document map and comes up with a text about H&M, closely neighboured to the annotated Boss articles. The text says that fashion discounter Hennes & Mauritz could improve its turnover by 12% in the last quarter, mainly due to its extraordinary turnover of casual wear, especially jeans and cotton jackets in Germany. The user selects this text in the map and finds that it is also contextualised by "Interest – Leisure and Casual" in the domain ontology. He is interested in other documents that report on casual wear and may indicate trends in this sector.

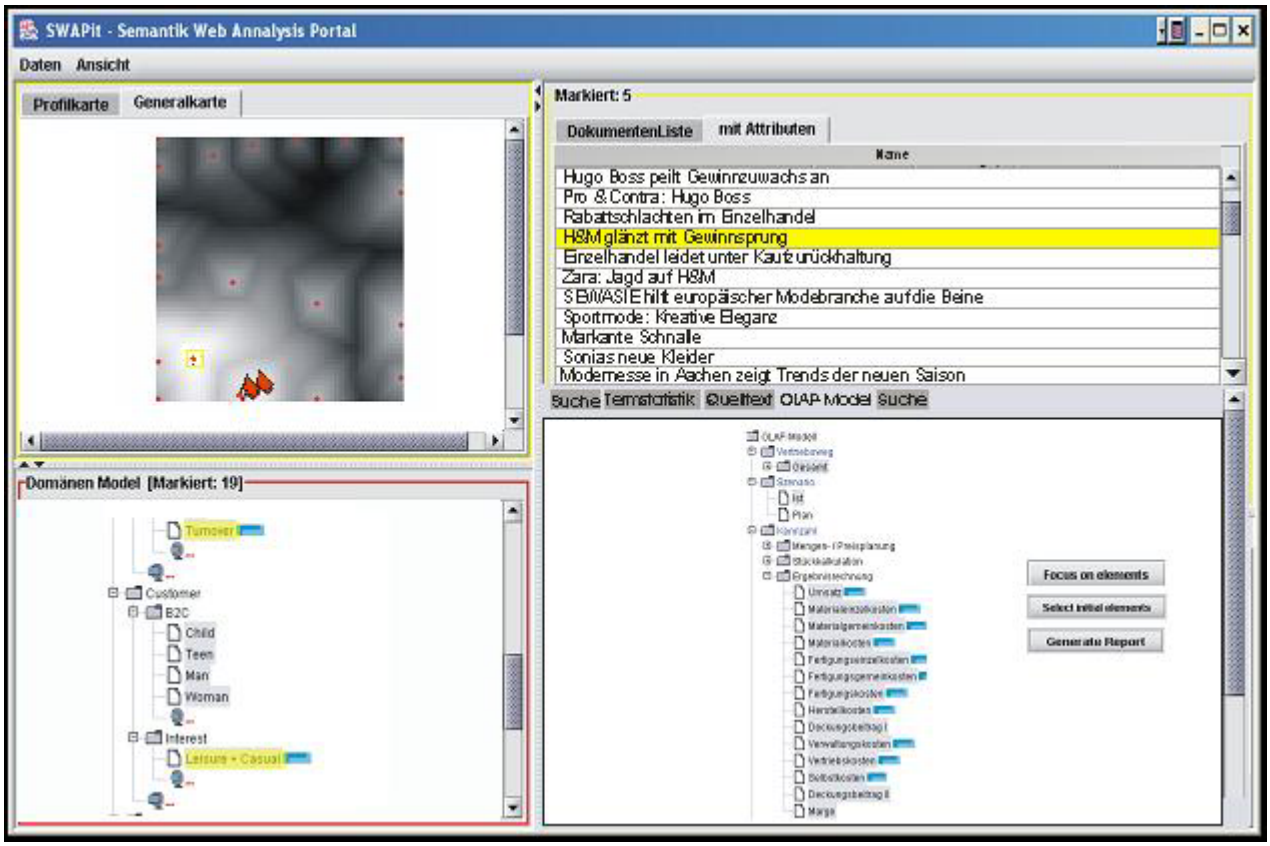


Figure 12: Exploring similar documents in the document map

The user clicks on the "Interest – Leisure and Casual" category in the domain ontology tree in order to explore other documents dealing with this issue understanding how these documents are related in the document map (highlighted in red, Figure 13). He finds that some documents similar to the H&M text deal with this topic, but also text from other clusters in the map are affected, among them a document on a fashion fair, pointing out future trends.

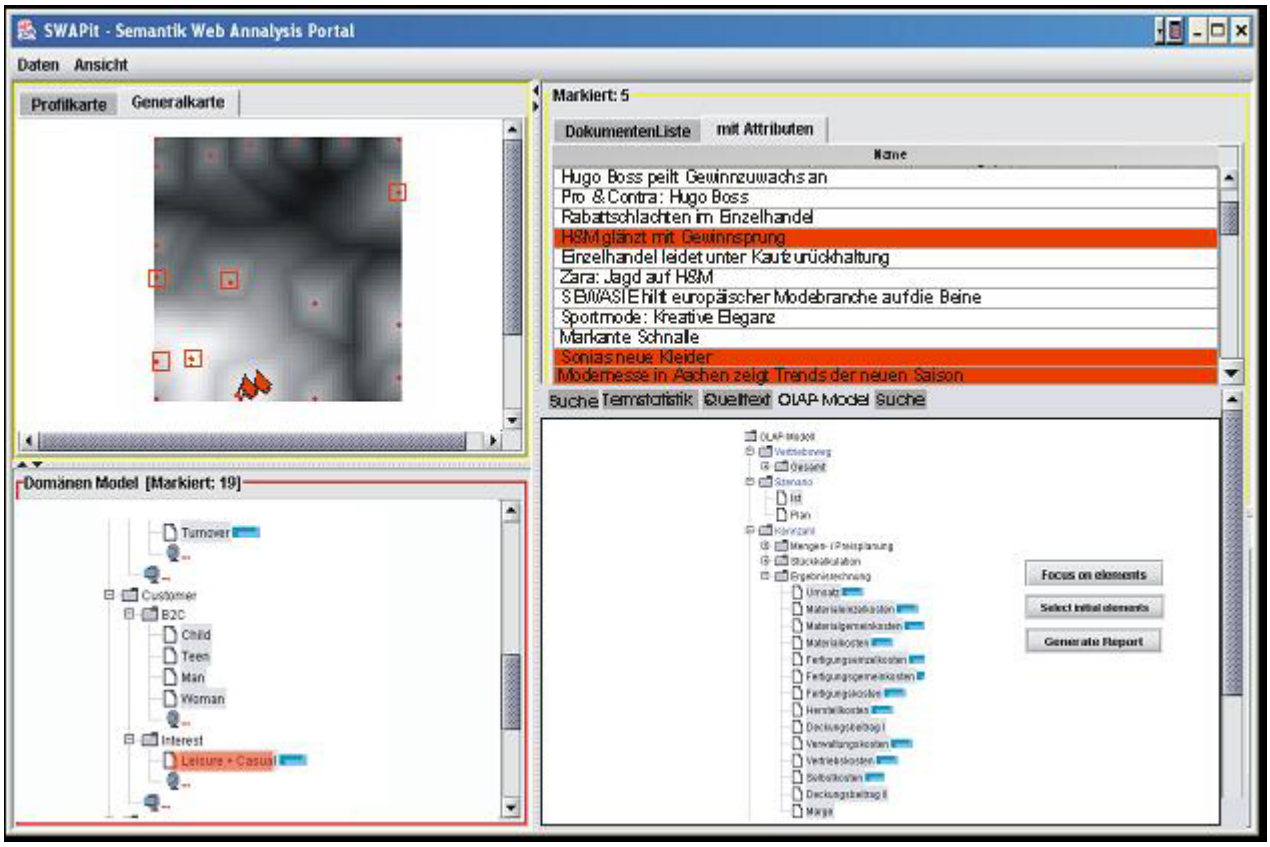


Figure 13: Selection in the domain ontology tree

7.3. Turning Back: Understanding Relationships to the OLAP Model

The user is satisfied with the information he collected. He understood that competitors are successful in particular in the sector of leisure and casual wear. Furthermore he learned about trends in this area. The analyst now wants to go back to his company's internal key data in order to learn more about its performance in the 'casual' sector. He looks at both, the domain model tree and the OLAP model displayed by SWAPit, in order to see how the specific text model concept "Leisure and Casual" relates to the OLAP model. He finds that the H&M text is also contextualised by "Season – Winter 2002-2003/Summer 2003" in the domain model. He moves his mouse pointer to the "Leisure and Casual" concept in the domain ontology tree. When moving over this concept, the mapping elements in the OLAP model view are highlighted (Figure 14). Now he sees that the concept is related to a specific set of elements in the clothing ('Kleidung') dimension of his OLAP model. He decides to configure a new report with cothing and season as running dimensions while keeping the other elements fixed in order to examine the company's performance in selling casual wear.

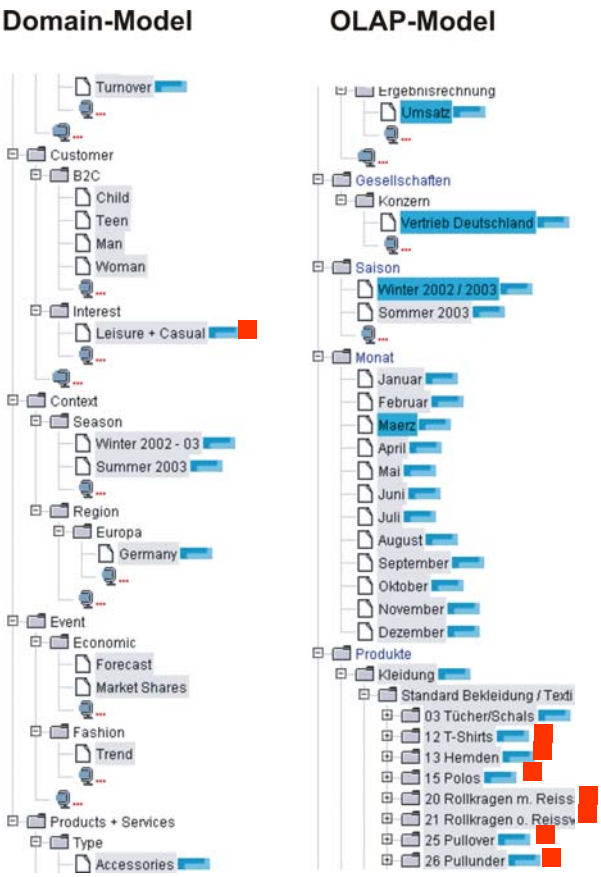

Domain-Model
OLAP-Model
Turnover
Customer
B2C
Child
Teen
Man
Woman
Interest
Leisure + Casual
Context
Season
Winter 2002 - 03
Summer 2003
Region
Europa
Germany
Event
Economic
Forecast
Market Shares
Fashion
Trend
Products + Services
Type
Accessories
Ergebnisrechnung
Umsatz
Gesellschaften
Konzern
Vertrieb Deutschland
Saison
Winter 2002 / 2003
Sommer 2003
Monat
Januar
Februar
Maerz
April
Mai
Juni
Juli
August
September
Oktober
November
Dezember
Produkte
Kleidung
Standard Bekleidung / Texti
03 Tücher/Schals
12 T-Shirts
13 Hemden
15 Polos
20 Rollkragen m. Reiss
21 Rollkragen o. Reiss
25 Pullover
26 Pullunder

Figure 14

8. References

[1] Allan, James, Leouski, Anton V., Swan, Russell C.: Interactive Cluster Visualisation for Information Retrieval. Tech. Rep. IR-116, Center for Intelligent Information Retrieval, University of Massachusetts, 1997

[2] Becks, A., C. Seeling, C. Engels et al.: Specification of the Architecture of the Monitoring Agent/Visualisation Component. SEWASIE Deliverable D4.1, November 2002

[3] Becks, A., C. Seeling, D. Frese, A. Hamm: Definition en evaluation of relevance measures for information matching. SEWASIE Deliverable D4.2, April 2003

[4] Becks, Andreas: Visual knowledge management with adaptable document maps. GMD research series 15, Sankt Augustin, 2001

[5] Codd, E. F.: Providing OLAP to User-Analysts: An IT Mandate, Hyperion Solutions Corporation, 1998

[6] Cody, W.F., J.T. Kreulen, V. Krishna, W.S. Spangler. The integration of business intelligence and knowledge management. IBM SYSTEMS JOURNAL, Vol.41, No.4, 2002

[7] Däßler R.: *Informationsvisualisierung – Stand, Kritik und Perspektiven*. In: Methoden / Strategien der Visualisierung in Medien, Wissenschaft und Kunst Wissenschaftlicher Verlag Trier (WVT), 1999.

[8] Franconi, E., S. Tessaris, T. Catarci, T. Di Mascio, G. Santucci, G. Vetere: Specification of the tool for the design, management and storage of the semantic information related to the data stores. SEWASIE Deliverable D6.2, February 2003

[9] Geroimenko V., C. Chen: Visualising the Semantic Web – XML-based Internet and Information Visualisation; Springer Verlag, 2002

[10] Hearst, M.A., Karadi, Ch. Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy. In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, Pennsylvania, USA, July, 1997

[11] Hearst, Marti A. TileBars: Visualisation of Term Distribution Information in Full Text Information Access. In: Proceedings of CHI 95, Denver, Colorado, 1995

[12] Jarke, M., Klemke, R., Nick, A.: Broker's Lounge - an Environment for Multi-Dimensional User-Adaptive Knowledge Management. *Proc. HICSS-34: 34th Hawaii International Conference on System Sciences*, Maui, Hawaii, January 2001

[13] Kay, J., B. Kummerfeld: User model based filtering and customisation of web pages. In Judy Kay and Bob Kummerfeld, editors, Workshop ”User Modelling for Information Filtering on the World Wide Web” held at Fifth International Conference on User Modeling, UM96, Kailua-Kona, Hawaii, Jan. 1996

[14] Klemke R.: *Modelling Context in Information Brokering Processes*. Doctoral Thesis, RWTH Aachen 2002

[15] Kohonen, T. *Self-organization of very large document collections: State of the art*. In Niklasson, L., Bodén, M., and Ziemke, T., editors, Proceedings of ICANN98, the 8th International Conference on Artificial Neural Networks, volume 1, pages 65-74. Springer, London, 1998.

[16] Lamping, J., Rao, R. and Pirolli, P.: *A Focus + Context Technique Based on Hyperbolic Geometry for Visualising Large Hierarchies*. ACM Conference on Human Factors in Computing Systems (CHI '95), Denver, Colorado, 1995.

[17] Lin, Xia. *Searching and Browsing on Map Displays*. In: Proceedings of the ASIS '95 Annual Conference, American Society for Information Science, October 9 – 12, Chicago, 1995.

[18] Mayer R., Freiberg U.: Workmanagement – eine wichtige Basis für Wissensmanagement; Jan. 2000; Wissensmanagement, Heft 1.

[19] Morse, E., M. Lewis, R. Korfhage, and K. Olsen. Evaluation of text, numeric and graphical presentations for information retrieval interfaces: User preference and task performance measures. Proc. of the IEEE International Conf. on Systems, Man, and Cybernetics, San Diego, CA, 1998, pp. 1026-1031.

[20] Nowell, Lucy T., France, Robert K., Hix, Deborah, Heath, Lenwood, Fox, Edward A. Visualising Search Results: Some Alternatives to Query-Document Similarity, In Proceedings of the 19th

Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zürich, Switzerland, August, 1996

[21] Shneiderman B. *The Eyes Have It – A Task by Data Type Taxonomy for Information Visualisations*. Proc. of the IEEE Symposium on Visual Languages, pp. 336-343, IEEE Computer Society, Boulder, Colorado, 1996

[22] Swan, Russel C., Allan, James. Improving Interactive Information Retrieval Effectiveness with 3-D Graphics, Tech. Rep. IR-100, Department of Computer Science, University of Massachusetts, Amherst, 1996

[23] Tavernari, Anna, Stefano Malatesta, Christoph Engels: Preliminary analysis of end-user requirements. SEWASIE Deliverable D8.1, September 2002

[24] Uhr, W., ed: *Externe Daten in Management Support Systemen*. Special Issue, Wirtschaftsinformatik, 41(5), pp. 403-457, 1999

[25] Veerasamy, A., Heikes, R. Effectiveness of a graphical display of retrieval results. In: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, Pennsylvania, USA, July, 1997

[26] Verity K2 Discovery Tier (Whitepaper: http://www.verity.com/pdf/ white_papers/)

[27] Vetterli, T., A. Vaduva, M. Staudt: Metadata Standards for Data Warehousing: Open Information Model vs. Common Warehouse Model. SIGMOD Record, 1999