



OKKAM – Enabling a Web Of Entities

Grant Agreement No. 215032

## D9.1: Mockup “Entity-centric organizational Knowledge Management”

---

Adrian Mocan, Wojciech Barczynski, Falk Brauer, Gregor Hackenbroich

|                                     |   |
|-------------------------------------|---|
| <b>Document Number</b>              | D9.1  |
| <b>Document Title</b>               | Mockup “Entity-centric organizational Knowledge Management” |
| <b>Version</b>                      | 1.0   |
| <b>Status</b>                       | Final   |
| <b>Work Package</b>                 | WP9   |
| <b>Deliverable Type</b>             | Demonstrator  |
| <b>Contractual Date of Delivery</b> | December 2008 (M12)   |
| <b>Actual Date of Delivery</b>      | 06 February 2009  |
| <b>Responsible Unit</b>             | SAP   |
| <b>Contributors</b>                 | SAP   |
| <b>Keyword List</b>                 |   |
| <b>Dissemination level</b>          | PU  |

## Change History

| Version | Date       | Status | Author (Company)   | Description   |
|---------|------------|--------|--------------------|---|
| 0.1     | 01/12/2008 |        | Adrian Mocan (SAP) | Creating the TOC  |
| 0.2     | 11/12/2008 |        | Adrian Mocan (SAP) | Adding to the section “Self support for customer”       |
| 0.3     | 04/02/2009 |        | Adrian Mocan (SAP) | Refining the “Self support for customer” section        |
| 0.4     | 05/02/2009 |        | Adrian Mocan (SAP) | Section 2.1 was filled                                  |
| 05      | 06/02/2009 |        | Adrian Mocan (SAP) | Section 3 was filled<br>Introduction/Conclusions filled |

## Executive Summary

This deliverable describes the current status of the two mock-ups developed in WP9, namely the *Self Support for Customers* and *Analytics of Unstructured Community Data*. These scenarios are described both from the user interaction and from the conceptual point of view, with a special emphasize on the role played by OKKAM in achieving the main targeted goals. The described features of these two mock-up applications are based on the requirements identified in “D1.2: Requirements Analysis for Entity-Centric Information Management in Corporate Environments” [2]. The current status and the ideas described in this document will constitute an input for the development of the actual prototypes that will eventually fully implement the features illustrated by these mock-ups.

## Table of Contents

|   |           |
|---|-----------|
| <b>1. INTRODUCTION.....</b>   | <b>5</b>  |
| <b>2. SELF SUPPORT FOR CUSTOMER.....</b>  | <b>6</b>  |
| 2.1. PRELIMINARIES .....  | 6         |
| 2.2. OKKAM USAGE SCENARIOS .....  | 8         |
| 2.2.1. <i>Support for entity disambiguation</i> .....                                 | 8         |
| 2.2.2. <i>OKKAM Search over OKKAMized Web resources</i> .....                         | 10        |
| 2.2.3. <i>OKKAM profile as additional information</i> .....                           | 11        |
| <b>3. ANALYTICS OF UNSTRUCTURED COMMUNITY DATA .....</b>                              | <b>12</b> |
| 3.1. PRELIMINARIES .....  | 12        |
| 3.2. OKKAM USAGE SCENARIOS .....  | 13        |
| 3.2.1. <i>Analysis of co-occurring entities in external information sources</i> ..... | 14        |
| 3.2.2. <i>OKKAM profiles as additional information</i> .....                          | 14        |
| <b>4. CONCLUSIONS .....</b>   | <b>16</b> |
| <b>REFERENCES.....</b>  | <b>17</b> |
| <b>APPENDIX 1.....</b>  | <b>18</b> |

# 1. Introduction

---

The development process of the WP9 applications has started with the implementation of mock-ups in order to allow the consortium as well as other third parties to provide feedback, suggestions and comments to the overall scenarios that are to be implemented. The *Self Support for Customer* and *Analytics of Unstructured Community Data* scenarios are only two examples of applications that can benefit from OKKAM and the entity centric approach it provides.

This document gives an overview of these two scenarios, emphasizing the OKKAM role and the benefits OKKAM brings to the overall goals of the two applications. The graphical interfaces and the interactions with OKKAM are analyzed in details. Additionally, throughout the whole document there are explicit references to the requirements derived as part of the D1.2 in order to show that the main conceptual and design decision were taken in accordance with these requirements. Appendix 1 lists the functional and the interaction with OKKMA infrastructure requirements and assigns each of them a number, such as they can be easily referred from the other section of the document. For a full description and explanation of these requirements please consult [2].

The document is structured as follows: Section 2 is dedicated to the *Self Support for Customers* scenario. It opens with an overall description of the implementation mainly from the user perspective and introduces the main elements of the graphical interface (Section 2.1). Sections 2.2, 2.3 and 2.4 describe the usage of the OKKAM platform in the context of this scenario. Section 3 presents the second use case, *Analytics of Unstructured Community Data*. In similar manner, some preliminaries are given followed by a description of OKKAM in this context. The deliverable finalizes with concluding remarks.

## 2. Self Support for Customer

---

In this section we give a brief overview of the *Self Support for Customer* scenario, emphasizing the role of the OKKAM infrastructure in supporting the overall showcase. We start by briefly introducing the main aspects of the showcase followed by a description of the main tasks addressed by using the OKKAM infrastructure.

### 2.1. Preliminaries

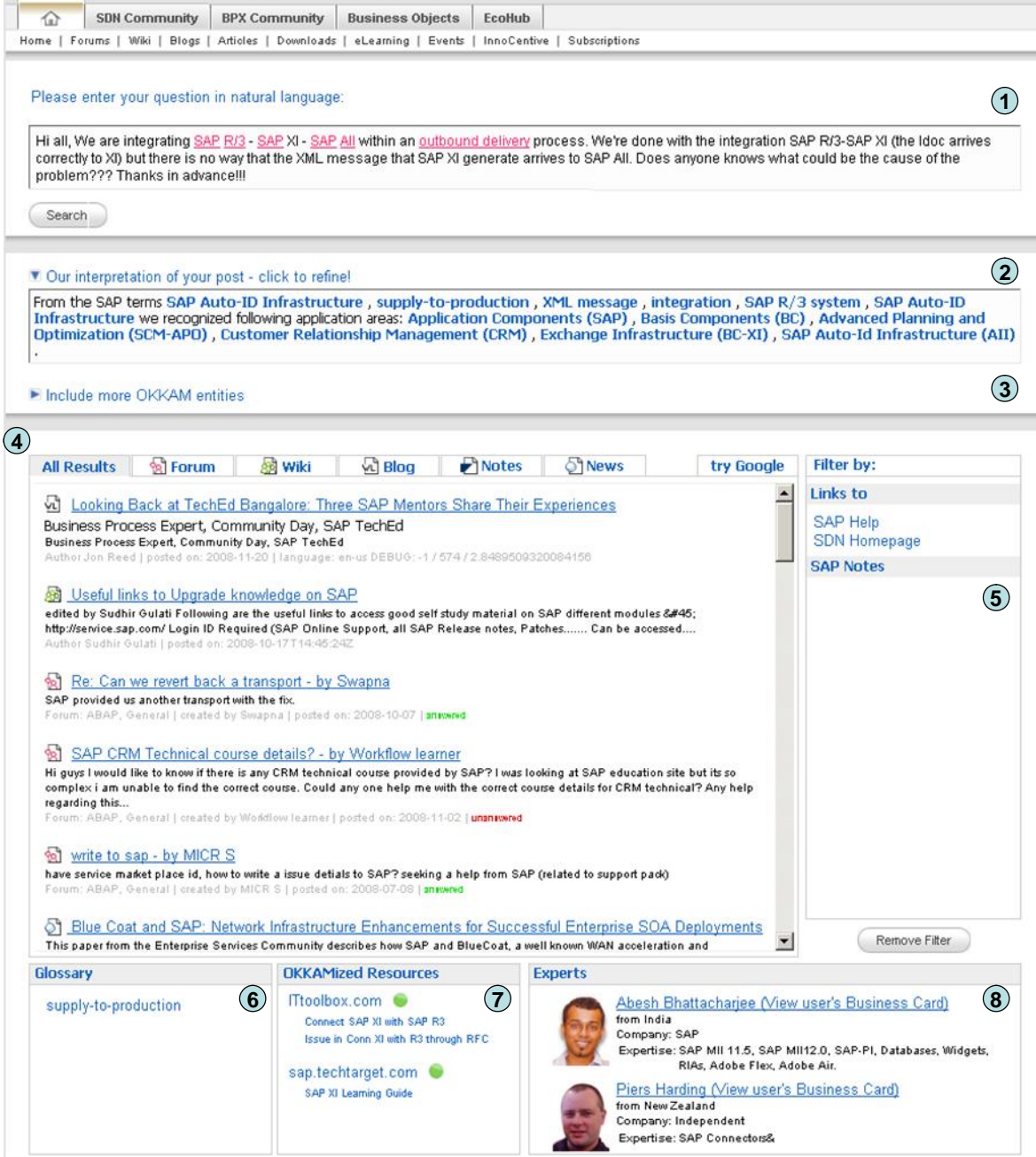
The *Self Support for Customer* scenario aims to provide advanced search functionality over unstructured data in the SAP Community Networks. The users (e.g. developers) can pose their queries in natural language form and the system provides an interpretation as a set of entities that are deemed to be relevant for the original query.

By applying the information extraction techniques developed as part of the SAP OKKAMizer, various entities are extracted from the unstructured text and matched with the entities already present in the corporate structured data. These matches are ranked by using a scoring model characterized by a *confidence* measure (indicating the degree of confidence in that match) and by an *informativeness* measure (estimating the significance of relations among the text fragments and entities in a certain context) [1]. These measures are aggregated in a *common score* that can be used to filter the final results, as specified in the functional requirements FR3, FR4 and FR5 (see Appendix 1).

The extracted entities can be removed or refined by the user. Based on the set of validated entities, the result set is created and classified based on the information source used. Besides this result set, several other resources are made available to the user in order to ensure that the appropriate answer is found: a glossary containing definitions of various terms, access to external information sources that might refer to the entities mentioned in the query and a list of experts in the area that could potentially contribute to solving the user problem.

The main components of the graphical interface shown in Figure 1 are described below (the numbers correspond to the numbered circles in Figure 1):

- (1) This panel is used by the user to post the natural language query.
- (2) It shows the interpretation of the original queries. The identified entities (terms and application areas or components) are shown in bold, blue text. The extracted entities are consistent with the functional requirement FR1 and with the OKKAM interaction specific requirements OIR1 and OIR2.
- (3) Additional entities, extracted by SAP-external OKKAMizer can be used as well for computing the result set. See Section 2.2.1 for more details.



The screenshot displays the OKKAM self-support customer scenario interface. At the top, there is a navigation bar with links for Home, Forums, Wiki, Blogs, Articles, Downloads, eLearning, Events, InnoCentive, and Subscriptions. Below this, a search bar prompts the user to enter a question in natural language. A question is entered: "Hi all, We are integrating SAP R/3 - SAP XI - SAP AII within an outbound delivery process. We're done with the integration SAP R/3-SAP XI (the Idoc arrives correctly to XI) but there is no way that the XML message that SAP XI generate arrives to SAP AII. Does anyone knows what could be the cause of the problem??? Thanks in advance!!!". Below the question, a search button is visible. The interface then shows an interpretation of the question, which lists various SAP terms and application areas. A search result page is displayed, showing a list of results categorized by source (Forum, Wiki, Blog, Notes, News). The results include forum posts, a Wiki page, and a blog post. A filter sidebar on the right allows filtering by source (SAP Help, SDN Homepage, SAP Notes). At the bottom, there are sections for Glossary, OKKAMized Resources, and Experts.

Figure 1: Self-Support Customer Scenario

- (4) The entities included in the interpretation are used to match annotated internal information sources and to prepare the result set. Panel 4 shows such a result set, classified based on the information source that contains that particular piece of information, e.g. Forum posts, Wiki pages, Blogs, Notes, etc.
- (5) The result set can be further filtered by using various facets. That is, one can choose to see only pages coming from the SAP Help or from the SDN Homepage.

- (6) It presents a Glossary for various terms relevant to the query.
- (7) This pane provides access to SAP external information sources that mention some or all of the entities included in the interpretation. For more details, see Section 2.2.2.
- (8) A list of experts intensively involved in one or more of the areas relevant to the original query, is presented to the user. This knowledge can be used to check other relevant information posted by these experts along the same topics.

## 2.2. OKKAM usage scenarios

This section gives an overview of the main OKKAM usage scenarios in the context of the Self Support for Customer showcase. Each of these scenarios uses the OKKAM infrastructure in order to provide a certain added value to the search facilities used inside of the SAP Community Network. That is, there are two main scenarios where the OKKAM infrastructure is used to improve the SAP customer self support facilities:

- Support for entity disambiguation;
- OKKAM Search over OKKAMized Web resources.

Additionally, the OKKAM infrastructure is used for an additional, less-critical task:

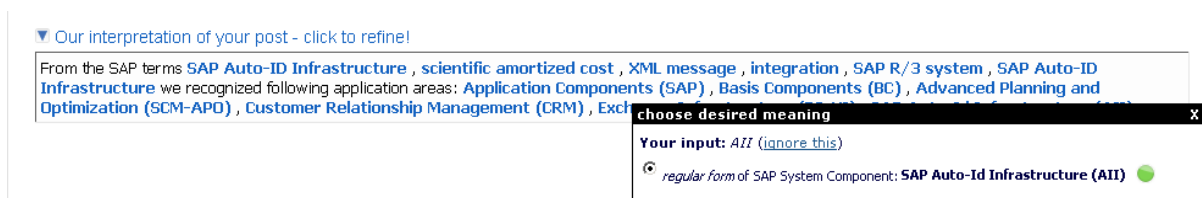
- Provide the OKKAM profile as additional information for the entity of interest.

Each of these scenarios is described in more details in the following sub-sections.

### 2.2.1. Support for entity disambiguation

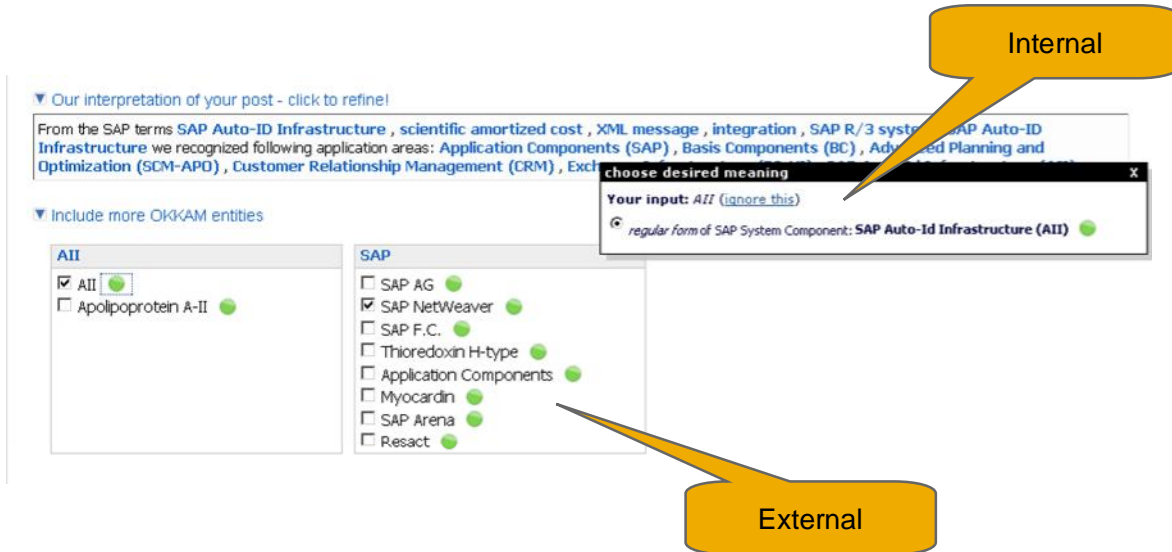
In our scenario, we use Information Extraction techniques to extract entities from the unstructured text (i.e. OKKAMization of the user query) and to link them with the existing entities in the corporate structure data [1] (functional requirement FR5). As the OKKAMization process has taken place beforehand for the corporate structured data (as specified by the requirements OIR7 and OIR9), the entities described in there have already an OKKAM Id associated. The OKKAMization of the corporate structured data can be run either as a one-time process or as an iterative process that compensates for the potential updates at the OKKAM nodes level.

The OKKAM infrastructure provides support for the disambiguation of the extracted entities (requirement OIR5). If the score of a particular match is higher than a certain threshold, the corresponding entity from the corporate structured data is included and presented in the interpretation (see Figure 2).



**Figure 2: Entities from the corporate data presented to the user for disambiguation**

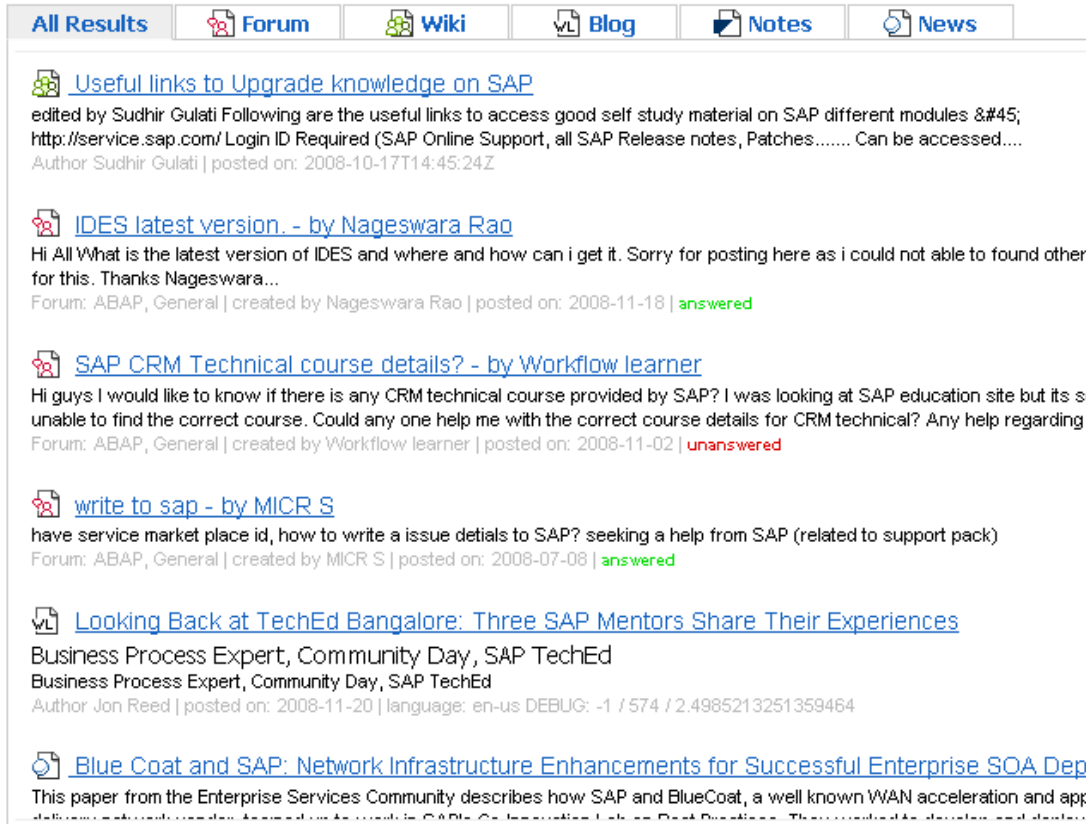
If the presented alternatives are not satisfactory, alternative entity definitions coming from external sources can be also presented to the user (see Figure 3). This is possible by using an external OKKAMizer<sup>1</sup> for the user query which will attempt to identify OKKAM entities extracted from external sources. Of course, if the internal entities included in the structured corporate data are registered with the OKKAM public nodes (which are being used by the external OKKAMizer), some of this internal entities might be retrieved as well.



**Figure 3: External entities presented to the user**

Based on the selected or ignored entities both in the internal and external set, the presented results are updated accordingly. The information sources directly affected by the internal entities identified in the interpretation are classified across several categories, such as Forum posts, Wiki pages, Blogs, Notes, etc. according with the company internal organization of the unstructured data sources (see Figure 4).

<sup>1</sup> At this stage, an OKKAMizer made available as a Web service by Expert Systems is used. The usage of the external OKKAMizer is in accordance with the requirements OIR4 and OIR11.

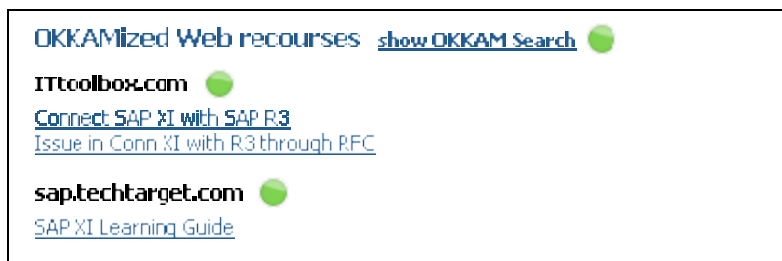


**Figure 4: Internal unstructured information sources delivered as part of the result set**

The external selected entities are presented to the user separately, as described in Section 2.2.2.

**2.2.2. OKKAM Search over OKKAMized Web resources**

As described in the previous section, an internal and an external OKKAMizer are used to extract the relevant entities from the user query. As such, each of the extracted entities (no matter if they have or not a correspondent in the corporate structured data) has an OKKAM Id associated. As a consequence, we can use the OKKAM Ids of the selected entities to perform a general search to find the external information sources that mention/refer to these entities altogether. We envision that a predefined set of “trusted” information sources (e.g. Web sites known to provide high quality information) will be searched in order to find the relevant information concerning several or all the entities of interest (see Figure 5).



**Figure 5: OKKAMized Web Resources**

Such a query exploits the full power of the OKKAM infrastructure and will be able to cross the corporate boundaries and find the answer the user is looking for. The answer set will be filtered based on a certain threshold and grouped over the information sources

### 2.2.3. OKKAM profile as additional information

After an interpretation is generated for the initial user query, a short description is provided for each of the recognized entities. Furthermore, if several candidates are identified, the user can directly contribute to the disambiguation of the extracted entity, by selecting the proper definition. As a consequence, the users have to be presented with sufficient information that can help them understand what the analyzed entity is about with a minimal effort. In this context, the OKKAM profile of the entity is the best option for fulfilling these needs: it contains only that information that can help in identifying the entity and it does not overwhelmed the user with too detailed information about that entity (see Figure 6).

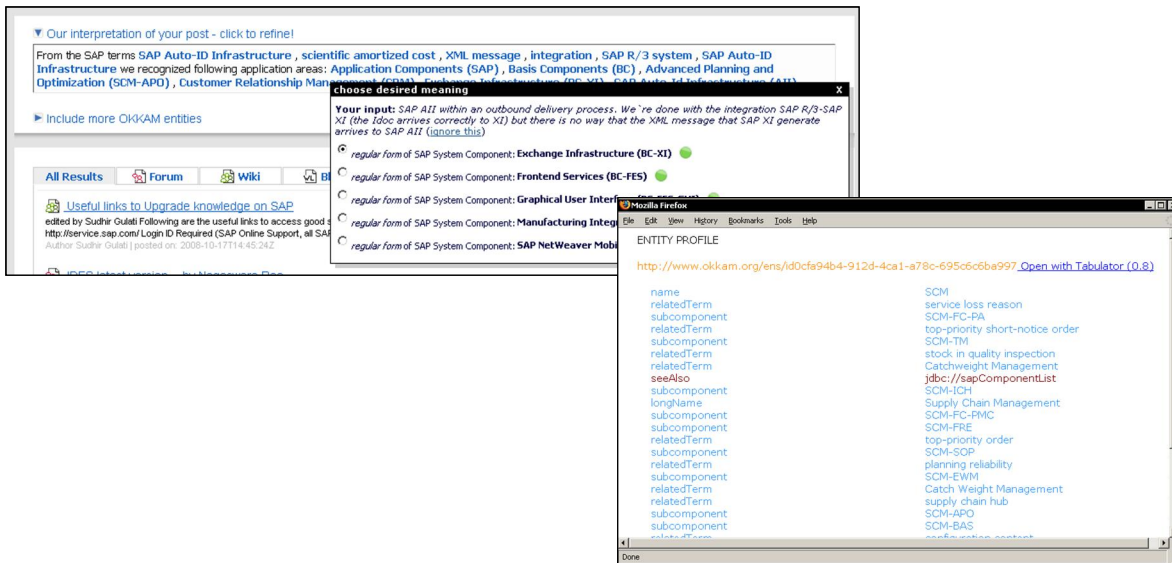


Figure 6: Using OKKAM profiles to provide the user with extra information about an entity

The usage of the OKKAM profile for providing additional information to the user is in concordance with the requirement OIR8.

## 3. Analytics of Unstructured Community Data

---

In this section we give a brief overview of the *Analytics of Unstructured Community Data* scenario, emphasizing the role of the OKKAM infrastructure in supporting the overall showcase. We start by briefly introducing the main aspects of the showcase followed by a description of the main tasks addressed by using the OKKAM infrastructure.

### 3.1. Preliminaries

The *Analytics of Unstructured Community Data* scenario aims at providing the business analyst with the means to obtain statistics and Business Intelligence (BI) insights over unstructured data. In this scenario, the unstructured data that is being analyzed comes from the user generated content available on the SAP Community Networks. Structured data is also being use as a point of reference or as an initiator of the analysis process.

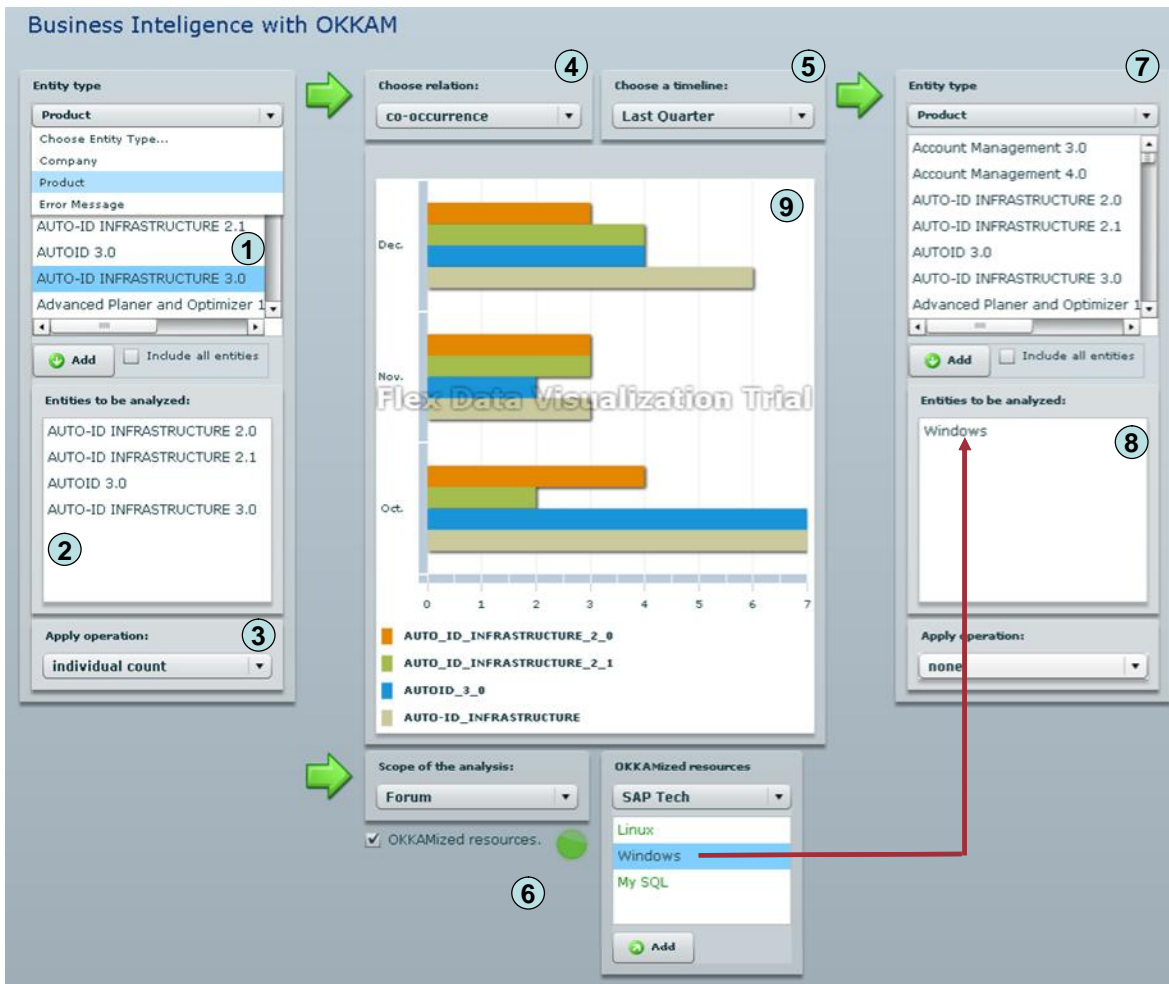
The BI analysis starts by browsing and selecting the entities of interest from the structured data available. The analysts can select the types of metrics they would like to use, as well as the scope of the analysis (e.g. the forum posts) and the information sources to be used. Next the users can select the type of relationship they are interested in and the timeline the analysis should span over. Finally, if the users are interested in co-occurrence relationships, they can browse and include more entities (with their corresponding metrics) in the overall analysis. Based on all these inputs and on the available data, a graphical chart will be created and displayed to the users.

The main components of the graphical interface shown in Figure 7 are described below (the numbers correspond to the numbered circles in Figure 7):

- (1) This panel allows the selection of entities from the structured data that are to be included in the analysis. These entities are classified in several categories such as Company, Product, Error Messages, etc.
- (2) The entities that have been selected in (1) are included in this panel. Only those entities displayed here will be eventually included in the analysis.
- (3) It allows the user to select and apply a metric to the selected entities. As the information source used in this scenario is unstructured data, the most relevant metric in this context is *count*.
- (4) It allows the user to select the type of relation that is to be identified and analyzed over the data in respect with the entities selected in (2). In the context of unstructured data, the most common relations are *occurrence* and *co-occurrence*. However, by using advanced extraction techniques more sophisticated relationships could be identified as well, such as *alternative of*, *competitor of*, etc.

Only, if the *co-occurrence* relationship is chosen the panels (7) and (8) are becoming visible.

- (5) The user can select the time span the BI analysis will range over. The most common time spans include weeks, years, quarters, etc.
- (6) This area allows the user to specify the granularity level at which the selected relationship will be sought for (forum level, post level, sentence level, etc.). Additionally, the user can also include external OKKAMized information sources in the analysis (see for more details Section 3.2.1).



**Figure 7: Analytics of Unstructured Community Data**

- (7) This panel is shown only if the *co-occurrence* relationship is selected. It allows the user to browse other entities included in the structured data and to add them to the analysis.
- (8) This panel is shown only if the *co-occurrence* relationship is selected. It contains the entities selected at (7) and any other entities extracted from the external information sources (see for more details Section 3.2.1).
- (9) It represents a graphical chart of the analysis. It is based on the parameters set at points (1) to (8) and on the available data. It is important to note that this data represents in fact annotations of the unstructured data from the SAP Community Network.

### 3.2. OKKAM usage scenarios

This section gives an overview of the main OKKAM usage scenarios in the context of the *Analytics of Unstructured Community Data* showcase. Each of these scenarios uses the OKKAM infrastructure in order to provide a certain added value to the BI analysis over the user-generated

content in the SAP Community Network. That is, there are two main scenarios where the OKKAM infrastructure is used to improve and extend the BI support over this unstructured data:

- Analysis of co-occurring entities in external information sources
- OKKAM profiles as additional information

### 3.2.1. Analysis of co-occurring entities in external information sources

As mentioned above, using the panel (6) the users can access external sources for obtaining additional information for their analysis. It is important to note that these external sources are OKKAMized information sources, which means that the entities they relate to are annotated with OKKAM Ids. The most interesting case for the BI analysis is if these external sources refer also to entities that are part of the company-internal structured data.

As a consequence, the first step is to search for the OKKAM Ids of the entities that have been included in panel (2). If such Ids are present (and depending of how the other parameters from (4), (5) and (6) are set) the related entities mentioned in conjunction with the entities from (2) are retrieved and presented to the user. The user can choose to include some or all of these entities in the analysis by adding them to panel (8) – this operation is illustrated by the red arrow in Figure 7.

By this process, the business analyst can extend the BI analysis across the corporate boundaries and reach to external sources for extra-information. We can easily imagine such external sources as being for example news Web sites (annotated with OKKAM Ids) and this extended analysis to be able to identify future trends or opportunities based on occurring events and situations.

### 3.2.2. OKKAM profiles as additional information

As in the previous scenario, the entity profiles from OKKAM can be used to provide information to the users and help them select the entities of interest (see Figure 8).

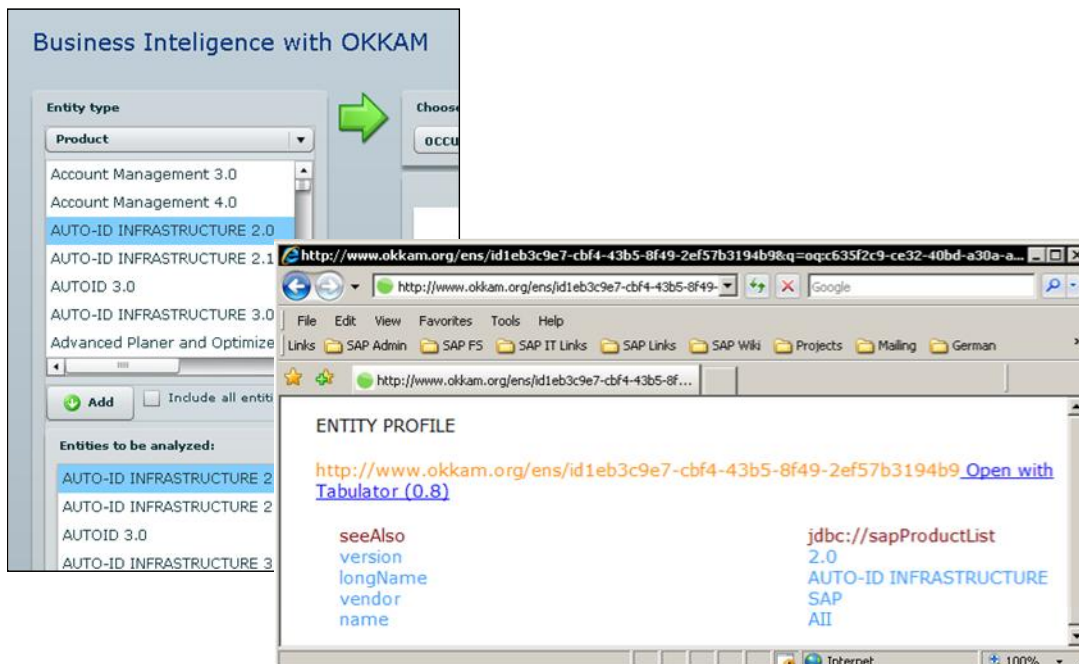


Figure 8: Using OKKAM profile to provide the user with extra information about an entity

This is useful when browsing the structured data or the entities retrieved from external sources in order to easier identify the entities relevant for certain types of analysis. The usage of the OKKAM profile for providing additional information to the user is in concordance with the requirement OIR8.

## 4. Conclusions

---

This deliverable has briefly presented the two scenarios developed as part of WP9. The user interfaces and the main conceptual ideas have been illustrated and the role played by the OKKAM infrastructure has been particularly emphasized. Additionally, several pointers have been provided to the requirements derived in [2] in order to justify the design decisions that have been taken.

As future work, the development in WP9 will continue with the actual implementation of the features that have been illustrated in the mock-ups and with the integration with the OKKAM platform. In fact, important steps have been already undertaken in this direction, such that several components of the *Self Support for Customer* application involve working prototypes and real data (for example the internal and the external OKKAMizers, retrieval of entity profiles based on the OKKAM Id, etc.).

## References

---

- [1] Falk Brauer, Marcus Schramm, Wojciech Barczynski, Alexander Loeser, and Hong-Hai Do: Robust recognition of complex entities in text exploiting enterprise data and NLP-techniques. In Proc. ICDIM '08. IEEE, 2008.
- [2] Wojciech Barczynski, Falk Brauer, Alexander Löser, Hong-Hai Do: D1.2 Requirements Analysis for Entity-Centric Information Management in Corporate Environments. OKKAM Deliverable. June, 2008

## Appendix 1

---

This Appendix lists and numbers the requirements that have been identified in D1.2 - Requirements Analysis for Entity-Centric Information Management in Corporate Environments [2].

### Functional Requirements:

- FR.1 Identification of domain-specific entities
- FR.2 Algebraic approach for Information Extraction
- FR.3 Support for ranking entities
- FR.4 Common metrics for extractors and data sources
- FR.5 Identification of given entities in unstructured data
- FR.6 Use of structured data to describe predefined entities
- FR.7 Disambiguation using context data

### Interaction with the Okkam infrastructure

- OIR.1 Extracting Product-related entities
- OIR.2 Extracting Terminology-related entities
- OIR.3 Extracting Knowledge-related entities
- OIR.4 Extracting Common entities
- OIR.5 Ambiguous semantics at world wide distributed sources
- OIR.6 Extraction knowledge kept is confidential
- OIR.7 Requesting OKKAM IDs
- OIR.8 Retrieving entity information
- OIR.9 Querying for similar entities
- OIR.10 Sending entity information
- OIR.11 Invoking publicly available Okkamizers

For the full description of these requirements please check [2]. The non-functional requirements have been not included since they are more relevant to the actual prototypes rather than to the mock-up implementations.