D5.4.2 – Integrated Dicode services (enhanced version)

Due date of deliverable: 31 December 2012
Actual submission date: 18 January 2013
Responsible Partner: UPM
Contributing Partners: UPM, CTI, NEO, FHG, UOL

Nature: ☒ Report ☒ Prototype ☐ Demonstrator ☐ Other

Dissemination Level:
☒ PU: Public
☐ PP: Restricted to other programme participants (including the Commission Services)
☐ RE: Restricted to a group specified by the consortium (including the Commission Services)
☐ CO: Confidential, only for members of the consortium (including the Commission Services)

Keyword List: Service Integration, Integration Framework, Integrated Workbench, User Interface, Web Application, Widgets, Service Oriented Architecture

The Dicode project (dicode-project.eu) is funded by the European Commission, Information Society and Media Directorate General, under the FP7 Cooperation programme (ICT/SO 4.3: Intelligent Information Management).
The Dicode Consortium

Computer Technology Institute & Press “Diophantus” (CTI) (coordinator), Greece

University of Leeds (UOL), UK

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (FHG), Germany

Universidad Politécnica de Madrid (UPM), Spain

Neofonie GmbH (NEO), Germany

Image Analysis Limited (IMA), UK

Biomedical Research Foundation, Academy of Athens (BRF), Greece

Publicis Frankfurt Zweigniederlassung der PWW GmbH (PUB), Germany
Deliverable manager

- Guillermo de la Calle, UPM

List of Contributors

- Guillermo de la Calle, UPM
- Víctor Maojo, UPM
- Miguel García-Remesal, UPM
- Eduardo Alonso, UPM
- Martha Rojas, UPM
- Manolis Tzagarakis, CTI
- Spyros Christodoulou, CTI
- Doris Maassen, NEO
- Jörg Kindermann, FHG
- Natalja Friesen, FHG
- Dhavalkumar Thakker, UOL
- Fan Yang-Turner, UOL

List of Evaluators

- Manolis Tzagarakis, CTI
- Jörg Kindermann, FHG

Summary

This deliverable reports on the advances carried out within tasks 5.3 (Dicode user interfaces) and 5.4 (Meaningful technical integration of Dicode services). It constitutes an enhanced version of Deliverable 5.4.1, presenting both updates of previous contents and new developments. It includes, among others, an update of the new features in the Dicode workbench, detailed explanation of the integration strategy followed in Dicode, and an exhaustive list of integrated suite of services. The intended audience of this document are end users of the Dicode services and therefore it has been elaborated as a user-oriented documentation emphasizing on how the services appear to and can be used by end users. Technical details have been avoided since they are intended for another audience and have been already reported in other deliverables (D3.2.2, D4.1.2 and D4.2.2).
Table of Contents

1 Introduction ........................................................................................................................................ 5
  1.1 Context & objectives..................................................................................................................... 5
  1.2 Structure of the deliverable........................................................................................................... 5

2 Dicode Workbench ............................................................................................................................ 6
  2.1 Introduction .................................................................................................................................. 6
  2.2 New features in the Dicode workbench......................................................................................... 6

3 Integration of services ....................................................................................................................... 8
  3.1 Introduction .................................................................................................................................. 8
  3.2 Integration approaches ................................................................................................................ 8
    3.2.1 Light integration ...................................................................................................................... 9
    3.2.2 Full integration ....................................................................................................................... 9

4 Suite of integrated services .............................................................................................................. 11
  4.1 Introduction .................................................................................................................................. 11
  4.2 Service description template ....................................................................................................... 11
  4.3 List of integrated services ........................................................................................................... 12
    4.3.1 Augmentor service .................................................................................................................. 13
    4.3.2 Collaboration service ............................................................................................................. 18
    4.3.3 Decision Making Support service ........................................................................................ 29
    4.3.4 Document Viewer service ................................................................................................... 40
    4.3.5 Emotion Detection Training service ...................................................................................... 43
    4.3.6 Entity Prominence service ..................................................................................................... 45
    4.3.7 Forum Summarization service ............................................................................................... 49
    4.3.8 Location of Twitter user service ........................................................................................... 51
    4.3.9 Opinion Mining service ......................................................................................................... 53
    4.3.10 Phrase Extraction Application service ................................................................................ 55
    4.3.11 Phrase Extraction Training service ...................................................................................... 58
    4.3.12 PubMed Mobile service ....................................................................................................... 60
    4.3.13 PubMed Mobile Extended service ....................................................................................... 64
    4.3.14 Recommender service ......................................................................................................... 68
    4.3.15 Recommender service for GEO datasets ........................................................................... 69
    4.3.16 Similarity Learning service .................................................................................................. 71
    4.3.17 Similarity Learning service for GEO datasets .................................................................... 72
    4.3.18 Storage service ..................................................................................................................... 74
    4.3.19 Subgroup Discovery service ................................................................................................. 78
    4.3.20 Subgroup Discovery service for functional interpretation of genomic data .................... 79
    4.3.21 Top Entities service .............................................................................................................. 82
    4.3.22 Top Hashtags on Twitter service (Keytrends service) ....................................................... 84
    4.3.23 Topic Detection service ....................................................................................................... 86

5 Conclusions & Future Work ............................................................................................................ 90

References .............................................................................................................................................. 91
1 Introduction

1.1 Context & objectives

This deliverable constitutes the enhanced version of the deliverable 5.4.1 describing the suite of integrated services in the Dicode project. This deliverable presents in a condense ways part of the work carried out in WP5 (Integrated Services & Innovative Work Methodologies). More specifically, it compiles the main outputs of task 5.3 (Dicode user interfaces) and task 5.4 (Meaningful technical integration of Dicode services).

Deliverable 5.4.2 is the second of a series of three deliverables reporting on the suite of services that have been developed and integrated in the Dicode project. The first deliverable of the series (D5.4.1) presented the initial version of the services developed during the first fifteen months of the project. The major objective of deliverable 5.4.2 is to present the new services developed between month 15 and month 28 within the Dicode project. Services will be described from the point of view of the end users. The overall idea is this documentation to function as the user manual of the developed Dicode services.

1.2 Structure of the deliverable

This deliverable follows a similar structure as the previous one. We have tried to avoid repetition of contents previously stated in the initial version. A short introduction has been included in each major section with two main objectives: first, to summarize the contents stated in the same section of the initial version and second, to introduce the new contents developed during the current reporting period.

Section 2 is devoted to the Dicode workbench. New features of the workbench—that already was presented in the initial version—will be presented. Section 3 is aimed to the integration of services. In this section, the final Dicode integration approach will be presented. Section 4 contains the description of the integrated services. Description of services is done according to a template created for this purpose. This template is also introduced in this section. Finally, section 5 presents some conclusions of the current reporting period about integration of services in Dicode. Moreover, some future work for the last year is sketched.

Section 4 (Dicode workbench user manual) and section 5 (Usage examples) of the initial version have been removed in the current version since their contents have not experienced substantial changes during this period.
2 Dicode Workbench

2.1 Introduction

The Dicode workbench is the key component produced in WP5. It is a widget-based web application enabling users to work together collaboratively and to use the Dicode services under a common interface. The Dicode workbench was introduced in the deliverable 5.4.1. The general description, site map and user roles detailed in that deliverable remain unchanged in this new version. New features of the Dicode workbench are detailed in the next section.

2.2 New features in the Dicode workbench

During the second year of the project, apart from fixing different bugs detected during testing, the workbench has evolved to improve usability and visualization of services. The current appearance of the workbench is presented in Figure 2.1.

![Dicode Workbench](image)

**Figure 2.1** – Current appearance of the Dicode workbench. Central part shows the widget that has been maximized (in the picture above, the collaboration widget appears in the central part of the workbench)

With respect to the workbench presented in deliverable D5.4.1, this new version includes the following new features:

1) **A new menu** appears on the top of the widgets. This menu allows users searching and adding new services to the workspace, save the current configuration, display and
modify the information of the workspace, access help, and close and exit the workspace. Search and Additional Info widgets have been removed from the workspace.

2) The Collaboration workspace, providing access to collaboration services, has been located inside a widget like the rest of the services.

3) All widgets can appear on any side of the workspace around the big central widget. No distinctions are made between Data mining widgets and other kind of widgets. In terms of their position, all of them are considered the same. Users can customize workspaces by reallocating the position of the widgets just dragging and dropping widgets from their original location to their final place.

4) All widgets can be maximized to occupy the big central area of the workbench. This way, services have more space to present their outputs. To move one widget to the center, users have to click in the arrows button located in the upper-right corner of the widget. Doing that, the widget exchanges its position with the widget in the center.

5) Widgets can be removed from the workspace clicking on the X button located in the upper-right corner of the widget. All widgets can be removed except the collaboration widget.

6) There is a beta version of the workbench using an elastic web design [1][2]. This kind of design enables applications to adapt dynamically their interfaces to use the full resolution of the client browser.
3 Integration of services

3.1 Introduction

This section in deliverable 5.4.1 presented the integration approach selected in Dicode project: widgets, registry of services and the single sign-on strategy. These parts remain unchanged and are not included in the current version. The section called Suite of services integrated has been moved to a separate section (Section 4).

In the current deliverable, this section reflects the advances and refinements carried out in the Dicode integration approach during the second year of the project. Section 3.2 presents the two types of integration defined for integrating services in the workbench.

3.2 Integration approaches

The Dicode workbench is envisioned as a mashup web application [3], allowing users to share resources under a common framework. Mashup applications usually consist on applications showing together different applications or components such as, for instance, iGoogle [4]. But traditionally those components neither share information nor are communicated between them at all.

In the Dicode project we are moving one step ahead. Users are enabled to move data from one widget to another just using the mouse and the drag-drop functionality developed for this purpose. The system architecture is designed to maintain as loose coupling among all integrated resources as possible.

During the design of the integration framework, we also took into consideration the recommendations raised by the reviewers in the first annual technical review. In particular, we focus on two comments: i) to consider the commonalities and differences among the three use cases to create a general framework with broader applicability, where use cases' scenarios are instances of such a framework, and ii) to consider the integration of existing (3rd party) tools. Finally, we designed an integration framework considering three major topics: flexibility, scalability and sustainability (see Figure 3.1).

![Figure 3.1 – Main features of the Dicode integration framework](image)

D5.4.2 (version 5): Page 8 of 91
Flexibility is needed to integrate any kind of service or application both internal and external to Dicode project. This framework should also consider both existing and future resources, implemented with any technology. Scalability is also a desirable feature to be able to expand the system over the time. In the Dicode integration framework, scalability is automatically given by the widget approach selected for the development of the workbench. The widget-based approach allows service providers maintaining their services running on their own machines. Thus, no big servers are needed to deploy the platform even if a large number of resource-demanding services are integrated. Services remain distributed and the computational load is also distributed. Finally, Sustainability is an essential feature mainly after the end of the project. The framework should allow integrating new services in an easy way, even for end-users, without any code modification. It can be found in the literature that sustainability along the time is a critical issue usually not considered by projects for their software products, making software useless when projects finish.

To deal with these issues, in the Dicode workbench two different approaches have been defined to integrate services and applications:

a) Light integration. It can also be named visual integration. This consists in the traditional mashup approach, i.e. services/applications are displayed together within the same web interface. No interactions happen between services, thus each service works as a standalone application.

b) Full integration. Not only services are displayed within the same application but data can be exchanged between them. Different technical mechanisms have been developed to communicate data among services. Web interfaces of services need to implement a set of functions to carry out properly the communication.

The Dicode workbench implements both integration methods at the same time. Service developers can select the level of integration desired for their services. The next two sections provide more details of these two approaches.

3.2.1 Light integration

We call Light integration the strategy followed by the mashup approach where different components are just displayed together within a common interface. To carry out this integration, service developers only need to develop the service, develop the web interface, deploy both elements in a web server and publish the service (URI to the web interface) in the Dicode workbench.

After these steps are successfully completed, services can be located and added by users to their shared workspaces.

3.2.2 Full integration

We call Full integration the integration approach that allows interactions between components integrated within the same common platform at both the graphical (aimed for end-users) and the programmatic (oriented to technical users) level. Interactions in the Dicode workbench are envisioned as events triggered when users move (drag with the mouse) items from one widget to another. Another extra feature of full integration is the definition of mechanisms to preserve state of services between invocations. Obviously, full integration covers all features of light integration.
We designed a loosely coupled architecture based on the idea of message passing interfaces [5] following a publish-subscribe design pattern [6]. In particular, we focused on the postMessage mechanism provided by HTML5 [7][8]. This mechanism allows applications running in different windows to communicate information (plain text) across different origins and domains. Although the content of the message only can be plain text, this is enough to communicate almost everything using, for instance, URIs or REST references.

As presented in Figure 3.2, the Dicode workbench acts as a message mediator between the different services or widgets. When the Dicode workbench detects that the user wants to move one element from one widget to another, it takes the reference from the origin source and send a message containing the reference to the target widget. Then, the target widget has to receive the message, interprets it and performs the actions associated with that message. Both reception and sending messages are optional for widgets (iframes), and it lies in the responsibility of service developers to incorporate them. Complete technical details for service developers will be gathered in the deliverable 5.1.2 Standards and guidelines for development (enhanced version) due in month 30. Meanwhile, those technical details are available in the wiki of the Dicode project¹.

![Figure 3.2 - Communication between widgets in the Dicode workbench during full integration.](image)

¹ [https://wiki.dicode-project.eu/display/DIC/Guidelines+extension+for+integration](https://wiki.dicode-project.eu/display/DIC/Guidelines+extension+for+integration)
4 Suite of integrated services

4.1 Introduction

In previous deliverable 5.4.1, the suite of integrated services was included as a subsection of section 3 Integration of services. In the current deliverable, we decided to establish a completely separate section devoted to this issue due to the importance of the suite of services within the Dicode project. In this section, services are described from the point of view of end-users. The objective is to provide end-users with a condensed (but complete) documentation of the services, to be able to understand and to use them in the workbench. Section 4.2 introduces the template that we have created to describe services. Finally, section 4.3 presents the suite of services.

4.2 Service description template

This section contains the template for describing services. This template has been created thinking in end-users, like a user manual or user help. Any reference to code, interfaces, classes or such kind of technical details have been avoided because they are out of the scope of this deliverable.

The service description template is as follows:

| Service name: |
| Description: |
| Type of integration: |
| Useful for: |
| Developed by: |
| Available at: |
| Additional information: |
| User interface description: |
| Example of use: |

*Note: all fields are mandatory, except Additional information.*

The meaning of each field in the service description template is as follows:

- **Service name:** indicates the name of the service
- **Description:** includes a short and concise textual description of the service; functionality, purpose, etc.

- **Type of integration:** indicates the type of integration implemented in the current version of the service regarding the Dicode workbench. Possible values are *Full integration*, *Light integration* or *Not integrated*.

- **Useful for:** lists the use case/s or user profile/s that the service is useful for. For instance, UCI or radiologists, biologists, etc.

- **Developed by:** indicates the Dicode partner/s that develops the service, using it’s short or full name.

- **Available at:** presents the URL where the service is available.

- **Additional information:** this field can be used to provide any extra and relevant information not covered by the other fields.

- **User interface description:** contains a detailed description of the user interface of the service, using screenshots and graphical representations when possible.

- **Example of use:** this field presents detailed examples of one or more scenarios where users can use the service, the actions that users should carry out and the windows/forms that will appear.

### 4.3 List of integrated services

In this section, the exhaustive list of the services integrated in the context of the Dicode project is presented. Services are listed in alphabetical order using the template presented in the previous section.
4.3.1 Augmentor service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Augmentor service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Augmentor service automatically tags the content (e.g. reports, documents, comments) with concepts from ontologies. It also allows searching and browsing content by terms and their relationship with other terms. It also discovers the most frequently used terms from the content (reports, comments, documents) and displays them as a tag cloud.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC2</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>UOL</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://imash.leeds.ac.uk/dicode/augmentor/augmentor.php">http://imash.leeds.ac.uk/dicode/augmentor/augmentor.php</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

Augmentor service is a web based tool to illustrate the utilisation of three semantic services:

- *Semantic Annotation service:* to tag content semantically, i.e. linking content to named entities
- *Semantic Query service:* to search (and to facilitate browsing of) semantically tagged content
- *Entity Summarisation service:* to summarise an entity

These three services are integrated within Augmentor. Augmentor is then integrated with the Dicode Workbench. Semantic Annotation service and semantic query service along with Augmentor service are described in the deliverable D5.3.1. The entity summarisation service is described in the deliverable D5.3.2.

The user interface of the Augmentor has a facility (Figure 1) to annotate textual content (e.g. reports).

![Figure 1. Annotation facility in Augmentor](image-url)
The annotation process invokes the semantic augmentation service, which in turn semantically tags the report and stores in a semantic triple store for browsing.

Augmentor provides an interface for searching and browsing of content. User can start their exploration using two options (Figure 2):

1. by typing search terms in the search text-box
2. by selecting popular terms from one of the tag clouds.

![Semantic search and tag cloud functionalities in Augmentor](image)

**Figure 2.** Semantic search and tag cloud functionalities in Augmentor

Figure 3 shows the typical page layout of an entity in the Augmentor.

![Information on “Inflammation” and tagged reports](image)

**Figure 3.** Exploration using ontological facts in Augmentor

Augmentor allows uni-focus exploration, i.e. there is a single point of focus (called “Focus
Entity”). For example, in Figure 3 the focus entity is “Inflammation”. Each page contains human readable description of an entity with a short abstract about entity and an image where available. Each page contains three facets about the focus entity (c.f. Figure 3).

The first facet contains facts about an entity. If facts are expressed in a triple (?s ?p ?o) then this facet contains the facts that have the focus entity as subject (?s). The second facet contains terms related to the focus entity. If facts are expressed in a triple (?s ?p ?o) then this facet contains the facts that have the focus entity as object (?o). The third facet contains content tagged with the focus entity using semantic augmentation service. Each of the items in all of the facets (e.g. facts, concepts, instances) are displayed with their labels (e.g. retrieved with rdfs:label, skos label properties or any custom- such as foaf:name).

In case where there is large list of facts about an entity, a shorten list of facts are displayed to the user (Figure 4). Such shortening is the result of entity summarisation service which creates a summary of an entity with few important facts.

![Figure 4. Displaying summary of facts in Augmentor](image)

**Example of use:**

Three radiologists are working on a multi-site drug trial for Rheumatoid Arthritis (RA). Each radiologist is generating reports from Dynamika- with comments on the analysis they have done. In such trials, the patients visiting can be with different type of inflammation and symptoms hence radiologists generally try to find out group of patients with similar RA symptoms (e.g. various types of inflammation, stiffness, sleep disorder). They would like to use a service that allows them finding reports on each of the symptoms or finding patients with a particular symptom (e.g. sleep disorder symptom).

For scenario mentioned above, and to address questions radiographers might have, they can start by typing “Rheumatoid arthritis” in the search box. That brings them to page informing about rheumatoid arthritis (Figure 5).
Figure 5. Displaying information related to Rheumatoid Arthritis in Augmentor

The page contains information on any reports that are tagged with “Rheumatoid Arthritis/RA” (by clicking on Expand button).

The page also contains summary of facts about RA from all the possible facts. As the radiologist are interested in the “signs and symptoms” related to RA, they click on “signs and symptoms” and arrive to a page about signs and symptoms.

On the page, signs and symptoms related to RA are listed. By clicking each of the symptoms and checking the reports for each of them radiologists can become aware of what type of symptoms are reported in the ongoing trial. For example, for knee inflammation, there were 5 reports (Figure 6).

Figure 6. Reports related to knee inflammation
On examination of each of these radiologists are able to conclude that there were only reported cases of “knee inflammation” and “wrist inflammation” and no reports for the rest of the symptoms (including sleep disorder that radiologists were interested to find out).
4.3.2 Collaboration service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Collaboration service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Dicode’s Collaboration service aims at offering innovative virtual workspaces which facilitate and augment collaboration towards sensemaking in data intensive settings.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Full integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1, UC2, UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>CTI</td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

Dicode’s Collaboration Service implements the view-based collaboration workspaces presented in detail in previous deliverables\(^2\). In particular, the service provides operations to create and manage workspaces and supports collaboration through the following views:

- **Mind-map view**, where the collaboration workspace is presented based on a spatial Paradigm;
- **Forum-view**, where the collaboration workspace is presented as a Web-based discussion forum.

Each view provides a different set of operations, which originate from the purpose of the view. The Collaboration Service provides its functionalities via a set of interfaces.

**Mind-map view**

In the mind-map view (Figure 1), the collaboration workspace is displayed as a mind map, where users can upload and interrelate diverse types of items. The mind-map view uses a spatial metaphor to organize items, which allows users to select and move around freely any item.

---

\(^2\) Deliverable D2.2: “The Dicode approach”

\(^3\) Dicode deliverable D4.2.1: The Dicode Decision Making Support Services (initial version)- [http://dicode-](http://dicode-)
Figure 1. Collaboration workspace operated in the mind-map view

Adding a collaboration item

Figure 2. Form to upload a new item onto the workspace, when operated in mind-map view

Adding a collaboration item on the collaboration workspace is possible through a form, which appears when double-clicking on the white area (i.e. an area that is not occupied by an item) of the collaboration workspace (Figure 2). Through this form, the user may set:
i) **The item type.** The mind-map view supports specific item types. The user may choose from the following types: idea, note, comment and service. Ideas stand for items that deserve further exploitation; they may correspond to an alternative solution of the issue under consideration and they usually trigger the evolution of collaboration. Notes are generally considered as items expressing one’s knowledge of the overall issue, an already asserted idea or note. Comments are items that usually express less strong statements and are uploaded to express some explanatory argument or point to some potentially useful information. Finally, service items enable users to represent, configure, execute and monitor data mining services from within the workspace, and allow the automatic appearance of their results into the workspace (as soon as the execution of the data mining service is completed).

ii) **The item’s content type.** The user may choose from simple HTML text, a URL or a file. By selecting the file content type, the user may upload a file on the collaboration space.

iii) **The item’s title** (any HTML formatted text). Text entered here will label the item’s icon in the workspace.

iv) **The item’s body** (any HTML formatted text), valid only when selecting content type text or URL. Content entered in this field will appear as the item’s body, visible when double clicking on the item in the workspace.

The collaboration item menu

When clicking on a collaboration item, a number of icons appear around the selected item, indicating a number of operations that can be performed on it. These icons are shown on the top and left sides of the item (Figure 3). The displayed options allow the following operations to be performed:

- **Creating a relation between two items (Figure 3 (a)).** Two collaboration items can be explicitly connected using directed edges. An example of such edge/connection is shown at (e) in Figure 3. A user may appropriately choose the width and colour of the edge to express a specific semantic relationship between two connected items. For instance, a thick red arrow connecting two documents denotes a strong contradiction. By clicking on a relation, the user may edit the relation and apart from resetting all the already mentioned attributes, he can set its width to state how strong the contradiction or agreement is. A legend can be also added on the edge drawn between the two objects stating further details, thus explicitly indicating the semantics of this interrelation (Figure 3 (f)).

- **Editing the item (Figure 3 (b)).** With the edit tool (Figure 3 (b)), the user may edit the attributes (item type, content type, title, text) set during the creation of the item and rate the item (a one-to-five stars rating scale is displayed). The deployed user and role model enables users to edit only the items that they have created.

- **Showing an item’s properties (Figure 3 (c)).** Clicking on the properties option allows displaying metadata for the selected item, such as its title, its creator, the date it was created, the user who last modified it, its modification date as well as its unique ID. Through the same form, the user may add dynamic (user-defined) attributes to the attributes of the item and tags.
Figure 3. Clicking on a collaboration item to show the item’s menu

- **Focusing on the neighbourhood of an object** (Figure 3 (d)). The neighbourhood option is used to focus on the neighbourhood of a collaboration item, allowing the display only of those items with which it is directly connected (via a relation). The aim is to allow users to focus only on directly connected items and not be distracted by others. Figure 4 shows the “neighbourhood view” of the collaboration item with title “Active Surveillance”. The selected item is depicted on the top (Figure 4 (a)) of the view, while the item’s neighbourhood are depicted beneath it with a coloured background depending on the type of relation the neighbours bear with the item (green for “in favour” (Figure 4 (b)), red for “against” (Figure 4 (c)), grey for “neutral”). The view also supports a number of operations. In particular, users may:
  - “like” or “dislike” an item (by using the buttons beneath the text of the item, Figure 4 (d-e))
  - update its title/content (Figure 4 (f))
  - create a new item (Figure 4 (g))

The total number of “likes” and “dislikes” received is depicted (Figure 4 (h)) for each item (textually and graphically through a coloured bar (Figure 4 (i))), while all users who “liked” or “disliked” the item are also viewable by hovering the mouse over their corresponding number. The user may also move to another item’s neighbourhood (clicking on the respective buttons appearing when mousing over the item’s text - by moving it up to the top level of the hierarchy) or choose to load all workspace items and their respective neighbourhoods.
Deleting an item

In order to delete an existing item from a collaboration workspace, users have to use the edit form (Figure 3 (b)) and click on the button with a red ‘x’, located at the bottom left part of the form. After successful completion, the item is removed from the collaboration workspace. Deleting an item will also delete any relationship that connects the item with other items in the collaboration workspace.

Configuring and executing an item of type “service”

When a new item of type “(data mining) service” is created, it has as icon an adjacent pair of gears appears on the workspace (Figure 5). Grey colored gears indicate that the service has not been executed. Through this item, a (data mining) service can be configured and executed. As in the case of any other item, by selecting the item of this type a number of icons appear indicating the operations that can be performed on it. The green triangle appearing in the set of available operations (Figure 5 (a)), allows users to configure and execute a (data mining) service.
Figure 5. Item of type “(data mining) service” which can be configured and executed.

By selecting this option, a dialog box appears allowing users to specify which (data mining) service to execute. Currently, the following data mining services can be configured and executed: i) Subgroup Discovery Service, ii) Phrase Extraction Service, iii) Similarity Discovery Service and iv) Recommender Service. After selecting the desired service, a form appears to specify its input parameters. As each service requires different parameters, a different form appears for each service. By completing the required inputs and pressing the “Go” button, the service starts its execution. When a service has been successfully started its execution, the colour of the gears in the mind map view changes to green. A green gear colour indicates that the service is still executing. When the service finishes execution, the gear colour change to orange and the results of the service appear on the collaboration workspace as separate items connected via edges to the service item.

Grouping collaboration items using adornments

Two or more collaboration items may be grouped together using a particular type of grouping mechanism called adornment (Figure 3 (g)). More specifically, adornments (depicted as coloured rectangles enclosing items) can be used to group together related objects. Such a grouping may be useful in aggregating items that are related to a particular issue or alternative solution (i.e. an answer/solution to the problem under consideration in the specific collaboration workspace); in such a way, alternatives may be better identified. The colour and title of these adornments are user-defined. Moreover, these adornments may share one or more documents (overlapping adornments). To create an adornment, users have to drag the mouse - while holding down the left mouse button and pressing the control key (ctrl) – and surround with the rubber band the items they wish to group. By releasing the mouse button, a new adornment will appear grouping the contained items.

The collaboration workspace menu

The collaboration workspace menu (viewable by pressing on the button “Menu” located on the bottom left corner of the collaboration workspace) offers access to functionalities which allow:

- Executing a number of multi criteria decision making algorithms (Figure 6 (a)) to find out which is the best answer to the problem under consideration based on the current state of deliberation (see details in the description of the Decision Making Support Services).
Calculating workspace statistics (Figure 6 (b)), by displaying the “workspace analytics view” of a collaboration workspace. Figure 7 shows the analytics view of a collaboration workspace. Its main purpose is to quantify activities and related items within workspaces, and display them to users in order to facilitate their understanding and offer insights on the evolution of a collaboration. This view is especially useful when collaboration workspaces grow large. In particular, the information shown in the analytics view include:

- the number of collaboration items each user has contributed on the workspace as percentage of the total items posted (users with top 10 contributions appear) (Figure 7 (a)).
- the total number of the workspace collaboration items over time (starting from the date the first object was created until the “present” day) (Figure 7 (b)).
- a bar chart with the number of “likes” received per collaboration item (collaboration items that received the top 20 numbers of “likes” are shown) (Figure 7 (c)).
- a bar chart with the number of “dislikes” received per collaboration item (collaboration items that received the top 20 numbers of “dislikes” are shown in this chart) (Figure 7 (d)).
- a bar chart with the “hottest” collaboration objects. The term “hottest” refers to the number of relationships in which the collaboration item is involved in the “mind-map view” (Figure 7 (e)).
- a bar chart with the score that each collaboration item receives. The score of each collaboration item is calculated as a weighted sum of four sub-scores: i) the difference between the number of “likes” and “dislikes” received, ii) the item’s creator score, iii) the number of relationships the item is involved, and iv) the
item’s rating. The default weights used to balance these sub-scores are: 0.2 for the item’s “like” and “dislike” sub-score, 0.2 for the item’s creator sub-score, 0.35 for the item’s number of relationships and 0.25 for the item’s rating. The larger score a collaboration item has, the more important is perceived for the discussion. (Figure 7 (f)).

![Workspace analytics view of a workspace showing statistics of several activities in the workspace.](image)

- Create an adornment on the collaboration space (Figure 6 (c))
- Replaying the actions performed on the mind-map view (Figure 6 (e)). In order to better help a user understand the evolution of a workspace, e.g. since the user’s last login, a replay mechanism is available. This mechanism is especially useful for users visiting a workspace for a first time in order to join an ongoing collaboration. Pressing the Replay button results in a blank workspace canvas, representing the workspace immediately after the workspace creation. The Replay mechanism window pops up on the right hand corner through which the user may “re-execute” each action.
performed on the workspace starting from the creation of the workspace. A number of options are offered such as step by step execution as well as moving backward or forward in time.

Mini map

Every collaboration workspace is equipped with a mini map (appearing when clicking the button “Minimap” on the bottom right corner of the collaboration space - Figure 6(f)) providing a full overview of the entire workspace. Within the mini map, the user is able to easily locate areas of dense activity. Finally, the mini map can be used for navigation purposes, to directly shift one’s view on another area of the workspace.

Forum view

In the forum view (Figure 8), the collaboration workspace is presented as a traditional Web-based discussion forum. The user may post new messages or edit existing ones. Messages posted through the workspace’s forum view are also visible in the mind-map view (as items of the ‘note’ type).

Figure 8: The forum view of the collaboration workspace

Example of use:

In the following scenario, an urologist (John), a physician (George) and a biomedical
researcher (Jane) are discussing alternative treatment options for prostate cancer patient. To discuss the alternative options, John creates from within the Dicode Workbench a new collaboration workspace and shares it with George and Jane.

John decides to use the collaboration workspace in mind-map view. To share his preferred way to treat the patient, he double clicks on the white area of the mind-map view which results in displaying the form (shown in Figure 2) to upload an item onto the collaboration workspace. As John prefers “Active Surveillance” as a solution to the issue being discussed, he chooses to upload an item of type idea by selecting the appropriate type in Figure 2. In addition, he adds “Active Surveillance” in the form’s title field to summarize his message and explains his preference in the body section of the form. By clicking the green “ok” button in the bottom right area of the form, an item of type idea is created and appears on the collaboration workspace. The newly created item features the entered title below the icon and displays information related to the creator and the time created.

Jane sees the item uploaded by John but has some concerns with respect to the proposed treatment. She thinks that the “Active Surveillance” method requires close monitoring of the patient, which involves regular digital rectal exams, PSA tests and prostate biopsy and many patients are unable to follow such a strict schedule. They neglect to make these exams which may lead in discovering late the worsening of their condition. Jane uploads her opinion by using the item upload form (Figure 2), selecting to upload an item of type comment and entering as title “Requires close monitoring” and explaining her concern in the body section of the form. Once the item has been uploaded onto the collaboration space, she indicates that her contribution expresses some concerns with respect to John’s proposal by using the connection tool (Figure 3 (a)) and drawing a relationship between her item and John’s proposal. To indicate that the item she uploaded argues against the “Active Surveillance” proposal, she clicks on the arrow and changes the colour to red.

As Jane is in favour of another treatment option, namely Brachytherapy, she shares it by uploading an idea item into the collaboration workspace using the appropriate form (Figure 2). John, seeing the alternative treatment option proposed by Jane, disagrees with the proposal because he believes that with this option there is no post-treatment staging possible which is important for such patients. Therefore, he uploads an item of type comment indicating his concern.

George, who participates in the collaboration, accesses the workspace and can see all the proposed treatment options and comments. He can easily see which comments argue against or in favour of each possible approach (by looking at the colour of the relationships) and read further information about each uploaded item by double clicking on it and revealing its body. Furthermore, he can access the properties of each item by using the properties tool (Figure 3 (c)) to get information related to when the item has been last updated and by whom. He may also use the Neighbourhood view of an item (Figure 4) to clearly display the items in favour or against. The Neighbourhood view allows George to vote/like or dislike individual opinions (Figure 4 (d) and (e)) in order to express his support. Seeing that a third treatment option has not been proposed, he uploads into the workspace an item of type idea representing his preferred way of treatment: “Radical prostatectomy”. John and Jane see the solution proposed by George and express their opinions.

As the collaboration workspace contains three different treatment options (“Active
Surveillance”, “Brachytherapy” and “Radical Prostatectomy”), which are - in the same collaboration workspace - discussed independently, John attempts to clearly identify and separate these three approaches by grouping the related items. Therefore, he creates three adornments (Figure 3(g)) surrounding the items that refer to each treatment option. He enters as title of the adornment the treatment options discussed by the grouped items. As each adornment receives a different colour, it easier for all participants to identify the areas of discussion. Participants may also upload “service” items (Figure 5) and execute them in order to get extra insights for the issue being discussed. As the results of the service will appear automatically on the workspace as separate items, they can treat them as any other collaboration item by commenting on them in order to collectively interpret the results.

As the collaboration progresses, more and more items appear on the workspace making it difficult to understand which alternative treatment option has received the greatest attention. To address such concerns, they can make use of the mind-map to see areas with high concentration of collaboration items (Figure 6(f)) or deploy the Analytics view of the workspace (Figure 7) to get insights related to items that received most attention.

For participants that favour interacting using Web-based discussion forums, they can decide to switch to the forum-view of the workspace (Figure 8) and participate into the discussion from this interface. Each participant may decide individually to make such a switch. The message a user is uploading to the workspace using the forum-view is also available at the mind-map view.
4.3.3 Decision Making Support service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Decision Making Support service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Decision Making Support services build on a formalization of collaborative decision making contexts to intelligently support stakeholders in such activities. They exploit machine-interpretable knowledge and diverse reasoning mechanisms to aid stakeholders in cases where they need to conduct formal argumentative discourses for the elaboration of alternative solutions to the problem at hand.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Full integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1, UC2, UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>CTI</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://hodgkin.dia.fi.upm.es:8080/dicode/workspace.jsp?id=workspace_id&amp;nm=workspace_title">http://hodgkin.dia.fi.upm.es:8080/dicode/workspace.jsp?id=workspace_id&amp;nm=workspace_title</a> (workspace_id: id of the selected collaboration workspace, workspace_title: title of the selected collaboration workspace)</td>
</tr>
</tbody>
</table>

User interface description:

Decision Making Support Services aim to intelligently support stakeholders in decision making activities by enabling the use and exploitation of appropriate reasoning mechanisms. They can be viewed through the “formal view” and the “multi-criteria decision making view” of a Dicode’s collaborative workspace.

The “formal view” \[D4.2.1]\(^3\) of a collaboration workspace permits a limited set of discourse moves for a limited set of message types whose semantics is fixed and system defined. In addition, this view can be associated with reasoning algorithms that are able to calculate which proposed solution (i.e. alternative) is currently prevailing or which position has been defeated.

The “multi-criteria decision making view” \[D4.2.2]\(^4\) of a collaboration workspace is a read-only view; its main purpose is to further support the decision making process by considering the attributes of the collaboration items appearing in the “mind-map view”. The quality of the decision making process has been augmented by exploiting algorithms coming from the field of Multi-Criteria Decision Making (MCDM). Elementary concepts of these MCDM algorithms are the alternatives and the attributes that represent the different dimensions from which the alternatives may be viewed. The main objective of a MCDM algorithm is, based on the attributes of each alternative, to calculate a corresponding score for the alternatives under consideration (in that the alternative with the higher score is

---

4 Dicode deliverable D4.2.2: The Dicode Decision Making Support Services (enhanced version) - http://dicode-project.eu/sites/default/files/D4.2.2-DecisionMakingSupportServices(enhanced)-v6-EC.pdf
considered to be the best solution to the problem at hand). The output of the MCDM algorithms implemented in the context of Dicode is a ranked list of the alternatives with respect to their scores (as calculated by the algorithm).

In Dicode, an alternative is represented on the mind-map view by a collaboration item of a specific type (the default type “idea” is used; however, this may vary according to the use case or workspace under consideration). The four attributes/criteria used to evaluate the alternatives are (more details can be found in [D4.2.2]):

1. ** Likes/Dislikes.** The algebraic sum of an item’s Likes minus Dislikes.

2. ** Creator Rating.** Calculated as the sum of all Likes/Dislikes corresponding to the items the Creator has contributed on a workspace.

3. ** Relationships in favour/against.** The algebraic sum of an item’s “in favour” relations (depicted with green arrows in the mind-map view) minus the item’s “against” relations (depicted with red arrows in the mind-map view)

4. ** Item rating.** The total rating corresponding to the users’ preferences (a user expresses his preference on an object through an one-to-five star rating scale)

Three MCDM algorithms have been implemented, namely: the Weighted Sum Model (WSM), the Analytical Hierarchy Processing (AHP) and the Lexicographic Decision Making rule (LDM) (refer to [D4.2.2] for details on the theoretical background of the algorithms and their implementation in Dicode).

**Formal view**

![Formal view of Dicode](image)

*Figure 1. An instance of the formal view of the collaboration workspace*
The formal view of a collaboration workspace is accessible through the option “Argumentation view” of the Collaborative Workspace widget in the Dicode workbench (Figure 1). It is an interactive view, where formal argumentation discourse takes place. Items of this view include:

- the issue, which actually corresponds to the problem under consideration (coincides with the name of the collaboration workspace);
- the alternatives, which refer to the different choices a decision maker has (all the possible solutions to the problem under consideration);
- the positions that may support or refute another position or alternative;
- the preferences which weigh the importance of two positions.

The formal view depicts the items created in the mind-map view of the collaboration workspace in a hierarchical way. Items follow a tree structure where the root (issue) is the name of the collaboration workspace (Figure 1 (a)). Nodes of the second degree are the alternatives (e.g. Figure 1 (b) and (e)). All items linked to the alternatives in the mind-map view are parsed (like a genetic tree) to complete the tree structure. On the produced tree, the relationship parent-child may be either positive or negative (argument in favor or against). This information is extracted, through parsing the relationships in the mind-map view (this is performed by following some transformation rules, i.e. a green link refers to a position “in favor” (Figure 1 (c)), a red link refers to a position “against” (Figure 1 (d))).

Apart from the items that have been created in the mind-map view and are depicted in the formal view, the user may create new items and interrelations among them through a number of functionalities, including adding a new alternative, adding a position to support or object an alternative (or position) and add a preference to express the relative importance of a position over another.

An underlying reasoning mechanism is used to inform users about the most prominent alternative (this alternative appears in blue color - Figure 1 (e)). This mechanism takes as input the alternatives, positions and preferences of the formal view. Each time an element is added on the formal collaboration workspace, the underlying reasoning mechanism is triggered to calculate the most prominent alternative.

**Adding an alternative**

Adding an alternative on the tree structure is possible through the menu appearing when right clicking on the root of the tree and selecting “Add alternative”. In the form showing up (Figure 2), the user may insert descriptive input for the alternative. Such input concerns fields such as the alternative’s subject, related URL, comments and an attachment file (describing the alternative in greater detail).

When the user clicks on the “Submit Alternative” button, a new node (Figure 1(b)) is created in the tree structure of the formal view. The text of the node is the text submitted in the “Subject” field of the form shown in Figure 2.
Figure 2. Adding an alternative in the formal view

Adding a position

Similarly, to add a position for supporting/refuting an alternative, the user may right click on the respective alternative and select “Add position” from the menu that pops up. On the form displayed (Figure 3), he may insert information for fields concerning the position such as its subject, a related URL, comments, and attachment file. He may also select whether the position to be created is in favour or against the specific alternative.

Figure 3. Adding a position in the formal view

When the user clicks on the “Submit Position” button, a new node (Figure 1(c)) is created in the tree structure of the formal view, named under the text contained in the “Subject” field used in the form of Figure 3.
Apart from adding a position to support/defeat an alternative, a user may add a position to support/defeat another position (accessible through a similar menu appearing when right clicking on a position).

Adding preferences

While working in the formal view, users may also add preferences (by right clicking on a position) in order to express their opinion about the relative importance of one position over another (Figure 4); a position may be “more important than”, “less important than” or “equally important with” another position. On the top of the form, the already added preferences are displayed.

![Image of adding a preference](image)

**Figure 4. Adding a preference**

Multi-Criteria Decision Making view

The Multi-Criteria Decision Making view of the collaboration workspace is accessible through the mind-map view.
Figure 5. Accessing the Multi-Criteria Decision Making view (button (a))

Pressing the menu button (a) of Figure 5, the Multi-Criteria Decision Making view of the collaboration workspace appears (Figure 6). The user may distinguish three areas, one dedicated to each MCDM algorithm implemented (Figure 6 (a), Figure 6 (b), Figure 6 (c)).

Figure 6: The Multi-Criteria decision making view

For each algorithm implemented (WSM, AHP, LDM), the user has to set the required
algorithm parameters. The algorithm may be then executed and what is returned is an ordered list of the alternatives (the output of the algorithm) in descending score order. Upon the execution of the algorithm, the Dicode user may browse through the detailed results of the algorithm (to realize the reason why an alternative performs better than another one) or view the plot with the scores of the alternatives.

**Weighted Sum Model (WSM) algorithm**

Concerning the WSM algorithm (Figure 6 (a)), the user may click on the “Set Weights” button (Figure 6 (e)) to change the predefined weights of the four factors (Figure 7) and, then, calculate each alternative’s score and sub scores. The four factors have to sum 100%, otherwise the execution of the algorithm is not possible.

![Figure 7: Setting WSM weights](image)

By using button (f) of Fig. 6, the scores and sub scores calculated by the algorithm are shown (Figure 8; each sub score corresponds to one of the four factors), while clicking on the button (d) of Fig. 6 shows the plot of the corresponding results (Fig. 9).

![Figure 8: WSM detailed results with alternatives scores and sub scores](image)
The output of the algorithm is a list of alternatives (Fig. 10) in descending score order (based on their total scores).

**Figure 9: WSM alternatives scores and sub scores plot (each color corresponds to a sub score)**

**Figure 10: WSM alternatives list.**

**Analytic Hierarchy Processing (AHP)**

With respect to AHP (Figure 6 (b)), the user may click on the “Set Weights” button (Figure 6 (g)) to open a wizard (Figure 11) in order to perform all the basic steps required by the AHP (these steps concern input of all required values stating the relative importance among all pairs of criteria and alternatives).

Similar to the WSM case, the two remaining buttons on the right provide access to the scores and sub scores calculated by the algorithm and the plot of the corresponding results.
Lexicographic Decision Making rule (LDM)

LDM is a decision rule based on ranking the attributes (in the case of Dicode, the four factors already stated) of the decision making process in terms of their importance. In Figure 6(c), the user has to press the corresponding button (Figure 6 (j)) to rank the four attributes based on their importance (Figure 12). Calculating the rank of alternatives is based on the partial score of the most important attribute.

In a similar way, the buttons on the right provide access to the scores and sub scores calculated by the algorithm and the plot of the corresponding results.

Example of use:
A physician (George), an urologist (John) and a biomedical researcher (Jane) aim to investigate which is the best alternative treatment for the prostate cancer. They decide to use Dicode to reach consensus. In this direction, Jane is setting up a collaboration workspace. She shares it with the physician and the urologist and they all move to the formal view of the collaboration workspace (Figure 1 – upon the creation of the workspace, only the top issue, Figure 1 (a), appears).

John suggests that one of the best and most popular treatments for the prostate cancer is the active surveillance. He right clicks on the top issue, clicks the “Add new alternative” option from the menu and he uses the form “Add new alternative” (Figure 2) that appears to post his statement (which is then displayed on the tree structure of the formal view – Figure 1 (b)). Jane is not in favor of this option, because it requires close monitoring (regular digital rectal exams, PSA tests, and prostate biopsy) to monitor for signs of progression, so she adds her “against” position on the collaboration workspace by right clicking on the alternative and using the form “Add new position” (Figure 3). Her objection is depicted on the formal view (Figure 1 (d)). Contrary to Jane, George supports the John’s opinion, in the sense that active surveillance avoids site effects from radiation therapy or prostatectomy, so he creates an “in favor” position supporting the alternative suggested by John (through the form of Figure 3 – position added on the formal view, Figure 1 (c)). He contradicts to Jane’s opinion and posts an “against” position because, according to his experience, most patients are unreliable as many, or most of them, neglect to visit doctors.

Jane underpins that Brachytherapy has been also used to treat tumors in many body sites and this could be one option (she posts an alternative on the collaboration workspace using the form of Figure 2). One of its major advantages is that this procedure does not need hospitalization (she creates an “in favor” position supporting the alternative she added (form of Figure 3)) and, furthermore, there are no surgical risks involved. John is not convinced by her arguments as Brachytherapy requires close monitoring (creates an “against” position (using the form of Figure 3)), which may even include hospital visits. He is so convinced about the truth of his statement that adds a preference stating that his position is “more important than” the one Jane posted before (he adds a preference of type “more important than” by using the form of Figure 4). To support his consideration against the Brachytherapy, John denotes that there is no post-treatment staging information which has also to be considered as an important factor (he adds an “against” position (Figure 3)).

George argues that the best option, in his opinion, is the radical prostatectomy as it is quite common with very good results (he posts an alternative on the collaboration workspace (form of Figure 2)). John is in favor of this option (so he creates an “in favor” position (Figure 3)) as this solution is proven to reduce prostate cancer death rates. Moreover, the removed tissue allows accurate stating (adds an “in favor” position (Figure 3)), which is very important and the PSA levels may reliably predict the recurrence (posts an “in favor” position (Figure 3)). Jane doesn’t share their enthusiasm as, first of all, due to the surgery a certain amount of risk is involved (“against” position (Figure 3)). Apart from this, an erectile dysfunction is expected at the level of 30-50% in 5 years (with nerve preservation) which reduces the parent’s quality of life.

The total collaboration process described is depicted in Figure 1 and, according to the input provided by the three collaborators, the alternative “Radical prostatectomy” is the winning
one (marked with blue font color).

Having exploited the functionalities of the formal view, the above stakeholders have not reached a final conclusion concerning the best treatment for the prostate cancer. Jane suggests using the multi-criteria decision making view of the collaboration workspace, where a number of MCDM algorithms may help them reach a more acceptable decision. As they all agree, they move to the mind-map view of the collaboration workspace, express their Likeness/Dislikeness and rating preferences [D4.1.2] on the collaboration items they posted on the formal view and, then, they move to the multi-Criteria Decision making view (Figure 6).

Jane believes that among the three offered algorithms the one closest to their needs is the Weighted Sum Model, so she sets the respective parameters (Figure 7) and browses through the detailed results of the algorithm (Figure 8), the graphical representation of the alternatives scores (Figure 9) and the results list (Figure 10). George believes that the best option to be used is the AHP algorithm as it’s a very popular algorithm in the area of multi-criteria decision making and it allows the pair wise comparison of both the criteria and the alternatives. He initiates the wizard to set the corresponding AHP weights (Figure 11) and calculates the scoring of each alternative. John is in favor of using the lexicographic rule as, according to his opinion, no compensation should be allowed among the four criteria (i.e. the alternative with the best partial score for the most important criterion should be the winning one). He uses the form provided for setting the order of the four criteria (Figure 12) and calculates the score for each alternative. Having used all three MCDM algorithms, stakeholders compare the respective results (alternatives’ rankings). Through such a sensitivity analysis, they are in a better position to reach a final decision.

---

4.3.4 Document Viewer service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Document Viewer service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service enables users to visualize distinct kinds of files (Word, Excel and PowerPoint documents, PDFs, TXTs, and HTML files). For opening and visualizing a given document, the service provides two ways: i) Dragging a file from the Storage service or other compliant service to this service. ii) Opening a file stored in a local PC. The service is fully integrated with the workbench, so files can be dragged and dropped to this service.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Full integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1, UC2, UC3. All kind of users that need to open and visualize documents inside the workbench</td>
</tr>
<tr>
<td>Developed by:</td>
<td>UPM</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://hodgkin.dia.fi.upm.es/viewer/">http://hodgkin.dia.fi.upm.es/viewer/</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

User interface description:

The main interface of the Document Viewer Service in the workbench is as follows:

![Figure 1. Idle State Interface](image)

The interface is quite simple, and it has three different states:

a) **Idle State.** You will see this interface the first time you add the service to a workspace or whenever no document is open. See Figure 1.

b) **Document-visualizing State.** When there is an opened document, the interface will look as presented in Figure 2. It displays the document in approximately 80% of the area, and in the 20% remaining an option for closing the file will appear.
c) **Open Document State.** This interface will appear when clicking “Open Document”. It contains a simple form for choosing a file, an “Open” button for confirming the action and finally a “Cancel” button for cancelling the action.

---

**Example of use:**

**Scenario 1:** John shares a workspace WS1 in the Dicode workbench with different people. They are conducting an experiment on miRNA. John has noticed that someone has added a new file to the Storage service, and he wants to see what this file may be about or contains. He simply drags the file to the Document Viewer widget for displaying it. While the document opens a corresponding loading alert will appear, and when the document is fully loaded, it will be displayed in almost the entire service area.
**Scenario 2:** John enters to a workspace and wants to open a PDF document stored in his computer. He doesn’t have a PDF reader installed; hence he can use the Document Viewer service for visualizing it. The user will click “Open Document”, choose the file from his computer, and wait for the document to load.

*Figure 5. Opening a document*
4.3.5 Emotion Detection Training service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Emotion Detection Training service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>The Emotion Detection Training service is very much like the Phrase Extraction training service in its functionality. There is one important difference, however. This service does not need a list of phrases to learn. It only needs a list of seed terms (that may include regular expressions). It will use all phrases around matching seed terms for the training.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Additional information:</td>
<td>Not operational yet.</td>
</tr>
</tbody>
</table>

User interface description:

The interface is like that of the Phrase Extraction Training service with one exception. Instead of a list of training phrases the user must supply a list of seed terms that the service uses to compile a list of training phrases itself. This list is a result of the execution (together with the extraction model). Please note that the list may be edited and re-used to train an even better model with the Phrase Extraction Training service.

Example of use:

A user wants to train the system to detect phrases that express an emotion of annoyance or bother in German texts in the automotive domain. She uploads a ZIP archive of 200 postings from the German Web forum “Motortalk”. Then she enters the seed terms “stört” (annoys) and “nervt” (worries) and starts the system. She gets an extraction model and a list of phrases detected by the extraction model. She removes the phrases that are inappropriate, and continues with the Phrase Extraction Training and the Application services.

First she uploads the extraction model and another collection of 100 postings to the Phrase Extraction Application service. Then she edits the resulting list of extracted phrases, removing the inappropriate ones. After that she uploads the joined phrase lists and all the 300 posting texts to the Phrase Extraction Training service to train a new and better model.
## Emotion Detection Training

<table>
<thead>
<tr>
<th>Upload text collection ZIP archive</th>
<th>[Durchsuchen...]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed terms (separated by #):</td>
<td></td>
</tr>
<tr>
<td>Phrase Label:</td>
<td></td>
</tr>
<tr>
<td>Model ID:</td>
<td></td>
</tr>
<tr>
<td>Specify language:</td>
<td>○ English</td>
</tr>
<tr>
<td></td>
<td>○ German</td>
</tr>
<tr>
<td>Other options:</td>
<td>○ Save model locally (enter Path)</td>
</tr>
<tr>
<td></td>
<td>○ Extract and store new phrases (enter Path)</td>
</tr>
<tr>
<td></td>
<td>○ Display new phrases after completion</td>
</tr>
<tr>
<td></td>
<td>○ Show tag cloud after completion</td>
</tr>
<tr>
<td>Execution:</td>
<td>Go!</td>
</tr>
<tr>
<td>Cancel:</td>
<td>[X] Close Window</td>
</tr>
</tbody>
</table>

*Figure 1. Emotion detection user interface*
### 4.3.6 Entity Prominence service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Entity Prominence service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>The Entity Prominence service displays a prominence graph for proper names (Named Entities) like brands, persons or places over time. In the graph, the user can compare the prominence dynamics of different entities. The service uses the same statistical data as the Top Entities service described above.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC3</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>NEO</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://dicode-project.eu:34080/dicode-entity-prominence/pages/prominence-graph-widget.html">http://dicode-project.eu:34080/dicode-entity-prominence/pages/prominence-graph-widget.html</a></td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the Entity Prominence service in the workbench is as follows:

![Entity Prominence service user interface](image)

*Figure 1. Entity Prominence service user interface*

Initially, the widget is empty, because no entities have been selected yet. When the user starts typing, potential matches are displayed in a dropdown menu.
After the user selects “Mercedes-Benz”, a prominence graph is plotted.

Figure 2. Showing potential matches of entities

Figure 3. Prominence graph displayed after entity selection
In most cases the user will add more entities to the graph and compare the prominence counts for different brands like in the following example:

![Prominence Graph](image)

**Figure 4. Prominence graph for multiple entity selection**

Additionally, the user can apply a set of filters:

- **Timeframe**: Select the start and the end date of the time frame.
- **Display by**: Select the granularity of the data points (Day, Week, Month, Year)
- **Languages**: Currently available languages are German and English

Afterwards, the user can start to explore the peaks of the graph. When hovering over a peak with the mouse, a circle appears. A click on the circle opens a Pop-Up-Window with a Google search for the term and date.\(^6\)

---

\(^6\) At the time of writing, Dicode does not offers search for the articles analysed for the Entity Prominence service, hence we are not able to display the original news documents which resulted in the peak in question. Google is used as an approximation, as in most cases a huge peak relates to a major event concerning a brand which will lead to a high amount in Google search results.
Figure 5. Exploring peaks of the graph for one concrete entity

Example of use:

Alice (Social Media Analyst) wants to analyze the conversations about the brand Mercedes-Benz. She uses the Entity Prominence service to see the dynamics of the conversations. The prominence graph shows the number of contributions for the entity over time. Alice compares Mercedes-Benz with the mayor competitors, BMW and Audi. Finally, she uses the service to dive into single peaks. She analyses the search results of the referring date to understand the drivers of the discussion. Alice jumps into promising links and analyses an interesting blog in depth to find some surprising insights.
4.3.7 Forum Summarization service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Forum Summarization service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>The service discovers the most discussed topics in an Internet Forum (such as Statistics Analysis forum of StackExchange) based on the content of the discussions. The topics are grouped based on their relevance and 10 groups are presented in 10 topic clouds, where bigger fonts represent more discussed topics.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UCI, UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>UOL</td>
</tr>
<tr>
<td>Available at:</td>
<td>A sample lightweight front-end interface website has been developed in PHP to visualize the topic clouds and browse the discussions using the topics in the clouds. The interfaces can be accessed from the following URL: <a href="http://imash.leeds.ac.uk/dicode/wp4/Stats2-Forum/">http://imash.leeds.ac.uk/dicode/wp4/Stats2-Forum/</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

User interface description:

To provide a visual output of the Forum Summarization Service (FSS), we have chosen to present the clustered groups and identified topics in a list of “Topic Clouds” (tag/word cloud, a weighted word list) as this format is useful for quickly perceiving the most prominent terms and to determine its relative prominence. The output of FSS includes 10 groups of topics, each of which is presented in a topic cloud, showing the most frequent terms (topics) mentioned in the text content for that group. Each topic cloud has a title, which shows the most discussed topic in that group, other topics listed in the word cloud are presented in different font sizes, which reflect the number of discussions.

For example, in the Figure below, the four groups of topics are the output from the statistical analysis forum “Cross Validated”. The topic “Distribution” was extracted from 307 discussions and related topics are shown in this cloud, such as “normal”, “probability” and “sample” etc. By clicking on a topic in the topic cloud (group title or words in the group), users can see all discussions from the group that include this word. By clicking on the title of the question, users can read the full detail of the question: the question and its answers, if any.

This user interface has been integrated into the Dicode workbench following the guidelines provided in D5.1.1: “Standards and guidelines for development (initial version)”.

If the users want to get a gist of an Internet Forum, the service can provide the most discussed topics and their related topics.
Example of use:

Sarah (PhD student), James (Postdoctoral Researcher) and John (Professor, supervisor of Sarah and James) are three researchers from a Breast Cancer research institution. They have conducted some studies on a small sample-size gene-expression microarray breast cancer dataset. The analyzed result is not satisfactory but they believe that some extra datasets from public resources, such as GEO (Gene Expression Omnibus) with the same pathology characteristics can augment their sample size and allow them to identify some extra statistically significant genes.

When making their decision on sample size, they want to search in appropriate forums for extra evidence/arguments. Using Forum summarization service, James quickly checked a forum (such as "Cross-validated", a statistical analysis forum) and see if the topics discussed are relevant to their research.
### 4.3.8 Location of Twitter user service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Location of Twitter User service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service shows the country or – if available – the city from where the tweet was published. The service uses the well-known Google maps interface which offers zoom functionality and an alternative “satellite view”. At first glance the results of the service are quite surprising. Brazil and Turkey show the greatest number of localized tweets. Although Twitter is increasingly popular in those countries, most Twitter users still live in the USA. There is a technical reason for this strange effect. Twitter users use lots of different tools to publish their tweets. Some users log in on the Twitter homepage, but many users use third-party applications like mobile apps for writing their tweets. Only some of the applications transmit the user’s geo coordinates for each tweet published to Twitter. Obviously, Brazilian and Turkish users prefer those applications.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>NEO</td>
</tr>
</tbody>
</table>

#### User interface description:

The interface of the Location of Twitter users service in the workbench is as follows:

![Image of the Location of Twitter users service user interface]

**Figure 1. Location of Twitter users service user interface**

---

7 In some countries, like for example China, Twitter's service is blocked.
By default the Locations of Twitter Users service shows the day before the current date. The user can adjust the date by using the date picker in the upper left corner.

![Locations of Twitter users](image)

**Figure 2. Filtering locations of Twitter users by date**

**Example of use:**

Alice monitors social media for a customer who targets global markets. She is especially interested in emerging markets in the BRIC countries. During the last couple of month, she noticed a trend in Brazil: Twitter has become increasingly popular in all parts of the country. Just recently also Chile appeared on the map. If a new green circle pops up, Alice searches Twitter for tweets from this country. In Alice’s experience, a green circle often refers to a place with a very vivid in innovative Twitter community: Twitter users in these places enjoy using mobile apps which exploit the full functionality of Twitter’s service.
4.3.9 Opinion Mining service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Opinion Mining service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>The Opinion Mining service is in some sense a combination of the Phrase Extraction Application service and the Topic Detection service. It first detects phrases in a text collection. Then it builds a topic model on those texts that contain at least one detected phrase. In a last step the topic model outputs the most significant sentences assigned with each topic. Supposed that the phrase detection model was trained on an emotion, these sentences will contain opinions of the text authors that are related to the topic in question.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Additional information:</td>
<td>Not operational yet.</td>
</tr>
</tbody>
</table>

**User interface description:**

- **Text Collection Source:** The user has to specify the text collection source and the text language in the same way as described in the Topic Detection Service.
- **Model Type:** Please see the Phrase Extraction Application service.
- **Specify language:** Please see the Phrase Extraction Application service.
- **Other options:** The user can specify the number of topics that are processed. The topics are sorted according to their weight in the topic model, which is computed from the number of texts assigned to this topic. The user may also specify the number of typical sentences displayed for each topic.

**Example of use:**

A user wants to determine opinions of forum posters from the German forum “Motortalk”. She uploads 3000 posting texts of the forum as a ZIP file. She also uploads the phrase extraction model that detects the “annoyance” emotion (see example of use in the Emotion Detection Training service description).

The Opinion Mining service outputs an overview of the 10 most important topics with 3 representative phrases of each topic.
## Opinion Mining

**Text Collection Source:**
- [ ] upload ZIP archive
- [x] Query text harvester

**Model Type:**
- [ ] Automotive texts
- [ ] Laptop texts
- [ ] Detect positive phrases
- [ ] Detect negative phrases
- [ ] upload model

**Specify language:**
- [ ] English
- [x] German

**Other options:**
- Number of top topics: 10
- Number of phrases per topic: 3

**Execution:**
- Go!

**Cancel:** [X] Close Window

---

**Figure 1. Opinion mining user interface**
4.3.10 Phrase Extraction Application service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Phrase Extraction Application service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>The service applies a pre-computed phrase extraction model to a text collection. The result is either the text collection with highlighted phrases or a tag could computed from the phrases.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC3</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://tmserver.iais.fraunhofer.de:9090/PhraeExtractionApplicationService">http://tmserver.iais.fraunhofer.de:9090/PhraeExtractionApplicationService</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

User interface description:

The user has to specify the text collection source and the text language in the same way as described in the Topic Detection Service. For the model type he or she has several options. The user can apply one of four pre-trained models that are stored on the server and extract positive or negative phrases from the domains of automotive texts or consumer electronics (laptop reviews). Another possibility is to upload a new extraction model which has previously been created using the Phrase Extraction Training service (see below).

The parameters of the Phrase Extraction Application service are explained in detail here:

- **Text Collection Source:** See the Topic Detection Service. The feature of sub-directories does not apply.
- **Model Type:** There are four pre-trained models on the server which can be used to extract positive or negative phrases from the automotive domain or the consumer electronics domain. The user may also upload and apply his or her own pre-trained extraction model.
- **Specify Language:** The user must specify the language of the text collection in order to activate an appropriate language grammar component that helps to optimise the phrase extraction process.
- **Other options:** Tagging the “Display new phrases” field will display the text collection with highlighted phrases. Tagging the “Show tag cloud” field will display a tag cloud that is built based on all the detected phrases.
- **Execution:** start the service.
- **Cancel:** cancel the service.

Example of use:

The user uploaded a car review text in German and configured the service to extract positive phrases. The Highlighted text is shown in figure 2, and the tag cloud is shown in figure 3.
Phrase Extraction Application

![User Interface](image.png)

**Figure 1.** Phrase extraction application user interface


**Figure 2.** Positive phrases highlighted in a car review, written in the German language
Figure 3. A tag cloud produced from positive phrases of the car reviews. Tags are in German language. Some prominent tags include: Allows/gives opportunity ("gibt"), Driver ("Fahrer") or Curves ("Kurven").
### 4.3.11 Phrase Extraction Training service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Phrase Extraction Training service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>The service produces a phrase extraction model from a text collection and a list of phrases considered interesting by the user.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC3</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>FHG</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://tmserver.iais.fraunhofer.de:9090/PhraseExtractionTrainingService">http://tmserver.iais.fraunhofer.de:9090/PhraseExtractionTrainingService</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

#### User interface description:

The interface allows uploading the text collection and the list of phrases to be used for training the model. The user can also obtain a list of phrases that have been detected by the newly built extraction model in the text collection. In general this list is different from the one provided by the user, because the extraction model is capable of generalization. The new list can then be inspected by the user, and phrases that are not wanted can be eliminated. Then the training can be repeated with the same text collection and the new list as the training reference list.

The parameters of the Phrase Extraction Training service are explained in detail here:

- **Upload text collection ZIP archive:** (see the Topic Detection service) The feature of sub-directories does not apply. A minimal amount of training material must be supplied in order to achieve a good quality. Our experience suggests providing at least 500 sentences. These should not form one single document/text – the training requires several dozen texts at least.

- **Upload training phrases file:** The phrase list can contain any phrases that the user considers to be interesting in this context/domain (positive phrases for example). It is very important that the phrases are of medium length (i.e. a phrase should consist of a “grammatical” sub-component of a sentence like “noun-verb-adjective-noun” in the case of English), and that they exactly match occurrences in the training texts. If a phrase is not matched character-by-character in a training text, the model construction process cannot use it.

- **Model ID:** The user should label the resulting model with a unique ID to be able to identify it for later use (i.e. upload to the Phrase Extraction Application service).

- **Specify language:** The user must specify the language of the text collection in order to activate an appropriate language grammar component that helps to optimize the model building.

- **Other options:** The user may save the model locally for later use with the Phrase Extraction Application service. The list of phrases extracted by the model may also be saved locally. Tagging the “Display new phrases” field will display the text collection with
highlighted phrases (considering the size of the training collection, this will be very lengthy!). Tagging the “Show tag cloud” field will display a tag cloud that is built based on all the detected phrases.

- **Execution**: start the service.
- **Cancel**: cancel the service.

**Example of use:**

A user wants to train the system to detect positive phrases in English texts in the domain of sports (soccer). She uploads a ZIP archive of 200 postings from an English soccer Web forum. Previously she read the postings and extracted a list of 150 positive phrases which she also uploads. She gets an extraction model and a list of phrases detected by the extraction model. She removes the phrases that are inappropriate, and continues: she uploads the extraction model and another collection of 200 postings to the Phrase Extraction Application service. Then she edits the resulting list of extracted phrases, removing the inappropriate ones. After that she uploads the joined phrase lists and all the 400 posting texts to the Phrase Extraction Training service to train a new and better model.

**Phrase Extraction Training**

<table>
<thead>
<tr>
<th>Upload text collection ZIP archive</th>
<th>Durchsuchen...</th>
</tr>
</thead>
<tbody>
<tr>
<td>upload training phrases file</td>
<td>Durchsuchen...</td>
</tr>
<tr>
<td>Phrase Label</td>
<td></td>
</tr>
<tr>
<td>Model ID</td>
<td></td>
</tr>
</tbody>
</table>
| Specify language | ○ English
⊙ German |
| Other options | □ Save model locally (enter Path)
□ Extract and store new phrases (enter Path)
□ Display new phrases after completion
□ Show tag cloud after completion |
| Execution | Go! |
| Cancel | [X] Close Window |

*Figure 1. Phrase extraction training user interface*
### 4.3.12 PubMed Mobile service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>PubMed Mobile service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>This service allows users to make consults to the PubMed database. Users can search papers or abstracts related with topics and read them through the service interface.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Full integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC1, UC2. All kind of users that need to search papers related to the biomedical domain.</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>NCBI</td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>This service has been published and integrated by UPM into the workbench but it has been developed by the NCBI (third party).</td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the PubMed Mobile service in the workbench is as follows:

The interface of the PubMed Mobile service comprises two different parts: i) a search frame for the medical topics, and ii) The space under the search box where the results of the search and the abstracts of these searches are placed.

User’s searches results appear as follows:

![PubMed Mobile search interface](http://www.ncbi.nlm.nih.gov/m/pubmed/)

*Figure 1. PubMed search interface*
Figure 2. PubMed search results for “Breast Cancer”

In this picture (Figure 2) we can see the search box and the results that this search produces. The search can be launched by clicking in the ‘Search’ button or by pressing enter key.

Example of use:

John is conducting a research using one workspace and he needs to search for Rheumatoid Arthritis papers. He adds the PubMed Mobile service to his workspace. The new widget appears in one of free spaces for widgets. He moves the widget to the center of the window and search for Rheumatoid Arthritis (Figure 3).
He sees some paper that looks interesting. He clicks in the link and sees the abstract of the document and others papers related with the paper selected (Figure 4).

**Figure 3. PubMed search results for “Rheumatoid Arthritis”**

**Figure 4. Details of selected publication, which resulted from a PubMed search**
He sees relevant information that can use in his research.

In addition, he can search information in papers related with the one that has been useful, or he can search for another topic related with his investigation. This action can be performed typing the new topic in the search box.
4.3.13 PubMed Mobile Extended service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>PubMed Mobile Extended service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service allows users to make queries to the PubMed database. Users can search papers or abstracts related with topics and read them in the service interface. In addiction the service provides users with a history of queries performed to the PubMed database. This way one user can follow the consults made by other user that is searching for similar information.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Full integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1, UC2. All kind of users that need to search papers related with biomedical topics.</td>
</tr>
<tr>
<td>Developed by:</td>
<td>UPM</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://hodgkin.dia.fi.upm.es:8080/pubmedintegration2">http://hodgkin.dia.fi.upm.es:8080/pubmedintegration2</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>This service has been integrated by UPM but some parts have been developed by the NCBI (third party). UPM has implemented the history of searches using the existing PubMed Mobile service.</td>
</tr>
</tbody>
</table>

User interface description:

The interface of the PubMed Mobile service in the workbench is as follows:

![PubMed mobile extended search interface](image)

*Figure 1. PubMed mobile extended search interface*

The interface of the PubMed Mobile Extended service comprises three different parts: i) a search field to type the biomedical topics to search, ii) the space under the search box where the results of the search and the abstracts of these searches are placed, and iii) the history frame where users can see the history of queries made by all users to the PubMed database.
User searches look as follows:

![PubMed mobile extended search results for “MRI”](image)

In this figure, we can see the search box and the results that this search produces. The search can be launched by clicking in the ‘Search’ button or by pressing enter key.

The history of queries made by the user appears as follows (Figure 3):

![User’s PubMed query history](image)

*Figure 2. PubMed mobile extended search results for “MRI”*

*Figure 3. User’s PubMed query history*
In this figure, we can see the main functionality of the history of queries. Searches are shown by day. User that made the query is written between parentheses. The combo box allows users to filter the searches showed by one specific user. In this case the consults of all users are shown. The link located in the upper right corner is to navigate back in the service.

Queries in which any paper has been consulted are shown with a ‘+’ on the left. This indicates that users can consult the same abstracts that other users consulted previously.

Example of use:

John is conducting a research on Rheumatoid Arthritis. Alice has told him about her intensive research on that topic. John opens the workspace that both share for the research in Rheumatoid Arthritis. He goes to the PubMed Mobile Extended service and opens the search history. Filter searches made by Alice and observes the terms that she searched for. Click on each of them and see the results:

![PubMed search history](image)

**Figure 4. Searching by using the query history**

John looks at the relevant results, but he needs more help. He displays all history and checks if another user has made queries related to rheumatoid arthritis. He sees that the user *TrialSponsorTS* has done an exhaustive search on the subject. Then, he displays those searches and see that he has seen several abstracts:
Figure 5. Details of search history

John thinks this is interesting information and updates his investigation with the new topics found.
### 4.3.14 Recommender service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Recommender service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>This service allows querying the Dicode system for information items that are similar to those provided as input. The important prerequisite is a similarity model that is learned by using the Dicode Similarity Learning service. Given a particular information item the user is interested in (the “reference object”), the most similar items will be then recommended to the user. The service first computes a similarity between the item the user is interested in and each item among the candidate items. In the second step, the algorithm orders all items according to their similarity to the reference object and returns the most similar ones.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC2</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>FHG</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://tmserver.iais.fraunhofer.de:8080/SIService/services/sl/process">http://tmserver.iais.fraunhofer.de:8080/SIService/services/sl/process</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>User interface description:</strong></td>
<td>No user interface is provided.</td>
</tr>
<tr>
<td><strong>Example of use:</strong></td>
<td>The Recommender Service can be used as a backend for a recommendation system in a wide range of application domains. The feasibility of the service is demonstrated on the example of Recommender service for GEO datasets.</td>
</tr>
</tbody>
</table>
### 4.3.15 Recommender service for GEO datasets

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Recommender service for GEO Data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service intends to provide the user with relevant and interesting datasets from Gene Expression Omnibus – the largest public repository for high-throughput gene expression data. A dataset in the context of the GEO repository is an item that defines a set of related Samples considered to be part of a study and describes the overall study aim and design.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC2</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

#### User interface description:

The user can formulate his needs in two different ways: by defining a dataset of interest from GEO repository or by filling in some text field. In the first case he has to specify the datasets by its unique GEO accession number. In the second case the user defines a dataset he is interested in by simply entering appropriate keywords and phrases into the corresponding fields:

- Title
- Summary
- Overall Design
- Experiment type
- Platform title
- Technology
- Organism
- Number of samples

These are the fields that are used in GEO to characterize a dataset. Clicking on the button “Go!” starts the execution of the service. The similarity model pre-computed in the Similarity Learning service is applied to rank all data sets according to their similarity to the data set of interest.

The service returns a list of 5 datasets that best satisfy the criteria defined by the user.
**Recommendation of GEO datasets**

**Dataset of interest:**

- Select by GEO ID: 
- Or define by fields:
  - Title: 
  - Summary: 
  - Overall Design: 
  - Experiment Type: 
  - Platform Title: 
  - Technology: 
  - Organism: 
  - Number of Samples: 

**Execution:** Go!

---

**Example of use:**

Sarah has a set of genomic data which is too small to retrieve meaningful results. She believes that some extra datasets from public resources, such as GEO (Gene Expression Omnibus), can augment their sample size and allow improving analysis results. She launches the GEO recommender service to search for some extra datasets that are similar to the data she has. Sarah types in the request describing the data and the methodology applied. All qualified datasets are provided in a list.

---

*Figure 1. User interface of recommendation of GEO datasets*
4.3.16 Similiarity Learning service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Similarity Learning service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service aims to create a similarity model from the user preferences, which can be used for user specific recommendations. Given an information object, the service delivers a similarity model that is learned from user feedback - a set of object pairs labelled as “similar” or “dissimilar”.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC2</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://tmserver.iais.fraunhofer.de:8080/SIService/services/starttrain">http://tmserver.iais.fraunhofer.de:8080/SIService/services/starttrain</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

No user interface is provided.

**Example of use:**

This service presents a generic learning framework that is able to operate on a wide range of information items from different research fields. The service intends to be used as a backend for a recommendation system. We demonstrate this in the example of the Similarity Learning service for GEO datasets.
### 4.3.17 Similarity Learning service for GEO datasets

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Similarity Learning service for GEO datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>This service aims to create a similarity model for comparison of GEO datasets</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC1</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>FHG</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://tmserver.iais.fraunhofer.de:8080/SIService/SimilarityLearning.jsp">http://tmserver.iais.fraunhofer.de:8080/SIService/SimilarityLearning.jsp</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>The service is an instance of the general Similarity Learning service, which was adapted to learn similarity Model for comparison of GEO datasets.</td>
</tr>
</tbody>
</table>

**User interface description:**

The user feedback is obtained in an interactive similarity learning process, which starts with the following interface:

**Start Similarity Learning Training Process**

![Image](image_url)

*Figure 1. User interface for executing the Similarity Learning Training Process*

After pressing the button “Go!”, the user is shown iteratively 15 pairs of datasets and is asked to mark them as “similar”/ “dissimilar”/ “don’t know” (Figure 2). The service remembers the user feedback and builds a similarity model according to it.
Figure 2. Instance of interface asking user to rate the similarity of datasets

Example of use:

Sarah wants to use recommender service for GEO datasets, but would like to perform recommendation according to her own preferences. She first starts similarity learning service for GEO data sets process and gives her feedback in training process. Once the model is learned, Sarah can launch recommender service and retrieve data sets that are similar to the one she has regarding her personalized preferences.
4.3.18 Storage service

Although this service was introduced in the initial version of the deliverable, we decided to include it again to follow the same description template as the other services.

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Storage service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service allows users to upload files to the workbench and share them with other users. There are two options for sharing files: i) upload the file to be stored by the service, and ii) provide an URI where the file is available. The service is fully integrated with the workbench, so files can be dragged and dropped from/to the storage service.</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Full integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1, UC2, UC3. All kind of users that needs to store and share files.</td>
</tr>
<tr>
<td>Developed by:</td>
<td>UPM</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://hodgkin.dia.fi.upm.es:8080/StorageService">http://hodgkin.dia.fi.upm.es:8080/StorageService</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the Storage service in the workbench is as follows:

![Storage Service](image)

*Figure 1. Dicode storage service showing available files*

The interface of the Storage service is divided in two main regions: i) a menu in the upper side, and ii) a tree view containing links to the files stored in the service. Those links can be dragged from the Storage service to other services in the workbench.

The menu contains three options:
a) **Upload.** This option enables users to upload a new file. When clicked, a popup window appears allowing users to select the file, its location and metadata information about it. Following the form contained in the popup window is presented:

![Storage Service](image)

**Figure 2. User interface for uploading a new file into the repository**

This form presents the following field:

- **Name**, indicates the identification of the file in the Storage service.
- **Description**, contains a short description of the file.
- **File format**, establishes the format of the file according to DON concepts.
- **File contents**, qualifies the content of the file according to DON concepts.
- **File location**, indicates where the file will be stored. Users can select two options:
  - **In the cloud**, means that the file will be uploaded and stored by the Storage service.
  - **URI**, denotes that user will provide an URL where the file is available. Therefore, in this case, the Storage service will only store a reference to the file.

Once the file has been successfully uploaded, it will appear in the tree view.
**b) Configure.** This option presents the configuration form of the Storage service. At this moment, there are not parameters to configure the service.

c) **About.** This option shows information about the authoring and version of the service.

**Example of use:**

John shares a workspace in the Dicode workbench with Alice and Peter. They are conducting an experiment on miRNA. John has found a brand new publication about a novel method to detect miRNA sequences in scientific publications. He wants to share this paper with their colleagues, so he decides to use the Storage service of the Dicode workbench. He opens the shared workspace and click on the “Upload” option from the menu of the Storage service. A popup window appears in his browser and John fill the form as follows:

![Storage Service](image)

**Figure 3. Uploading a new file into the repository**

He has the paper in PDF format and prefers that the file is stored by the Storage service. Then, he clicked on the “Browse” button and selects the file from his hard disk. Finally, he clicks on the “Upload now” button and the file appears in the Storage service as follows:
Figure 4. Storage service showing the new file uploaded

Now the paper is available for Alice, Peter and John from the Dicode workbench.
4.3.19 Subgroup Discovery service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Subgroup Discovery (SD) service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>This service allows users to analyze their data using subgroup discovery algorithm. The algorithm detects novel knowledge given by subgroup patterns with regard to the property of interest. The algorithm includes a new mechanism for finding relevant subgroups, which is important to get non-redundant results.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC1</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>FHG</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://tmserver.iais.fraunhofer.de:8080/SgdService/services/process">http://tmserver.iais.fraunhofer.de:8080/SgdService/services/process</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

No user interface is provided.

**Example of use:**

The SD service is designed as a backend for any kind of tasks using subgroup discovery algorithm. The service is generic enough to be applied in a wide range of domains. Despite its generality, the service can be easily targeted to more specific tasks, such as analysis of genomic data. We demonstrate this in the example of a Subgroup Discovery Service for functional interpretation of gene data, which is an instance of the SD service.
4.3.20 Subgroup Discovery service for functional interpretation of genomic data

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Subgroup Discovery service for functional interpretation of genomic data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service allows functional interpretation of genomic data. The service is based on a subgroup discovery algorithm and automatically includes external knowledge databases such as Gene Ontology (GO).</td>
</tr>
<tr>
<td>Type of integration:</td>
<td>Light integration</td>
</tr>
<tr>
<td>Useful for:</td>
<td>UC1</td>
</tr>
<tr>
<td>Developed by:</td>
<td>FHG</td>
</tr>
<tr>
<td>Available at:</td>
<td><a href="http://tmserver.iais.fraunhofer.de:8080/SgdService/SubgroupDiscovery.jsp">http://tmserver.iais.fraunhofer.de:8080/SgdService/SubgroupDiscovery.jsp</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>The service is an instance of the subgroup discovery service which was adapted for the tasks of functional interpretation of genomic data</td>
</tr>
</tbody>
</table>

User interface description:

The interface of the Subgroup Discovery for functional interpretation of genomic data service in the workbench is as follows:

**Subgroup Discovery on Gene Data**

![Interface Image]

**Figure 1. User interface for Subgroup Discovery service**

The interface contains the following fields:

- **Select input file** is a file browser to upload a data set. Valid file formats are .csv or .txt. The input file has to contain at least 2 columns separated by semicolon. In the first column are gene symbols. The second column contains binary values indicating whether the corresponding gene is correlated with the property of interest. A short
example of the input file is presented below:

<table>
<thead>
<tr>
<th>GeneName;Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRPF1;0</td>
</tr>
<tr>
<td>SLC12A4;0</td>
</tr>
<tr>
<td>LOX;0</td>
</tr>
<tr>
<td>KCNJ8;0</td>
</tr>
<tr>
<td>RIN1;0</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>NKG7;0</td>
</tr>
<tr>
<td>DOK1;1</td>
</tr>
<tr>
<td>ROS1;0</td>
</tr>
</tbody>
</table>

- **Number of rules** defines how many subgroups will be shown to the user.
- **Use ontology** field specifies the category of Gene Ontology that will be used to describe genes. The following three categories of GO are available:
  - Biological process
  - Cellular component
  - Molecular function
- **Attribute to include** (optional) – the user can specify the gene ontology terms that have to occur in the output subgroups. Notice: be careful with this field, specifying a term that never occurs in the data results in no subgroups.
- **Attribute to exclude** (optional) – the user can exclude some terms from the output. The both parameters “Attribute to include” and “Attribute to exclude” are helpful to optimize the first results of subgroup discovery algorithm and to achieve better results.
- **Execute pressing the “Go!” button starts the execution of the algorithm.**

### Example of use:

Sarah wants to find the functional patterns in a dataset. She launches the Subgroup Discovery service and uploads a file containing a list of gene names and a list of binary values indicating expressed genes. The service produces a list of subgroups sharing similar biological properties. The results of the service are displayed in a collaborative workbench and can be discussed with other collaboration members. Each subgroup is represented as a single element connected to the data mining element.
Figure 2. Data mining element uploaded into the collaboration workspace (orange colored gears) after it has finished execution. Results of the data mining service are also shown.

In order to see subgroup the user can open single text element.

Figure 3. Showing the output of the data mining service ("Results") by double clicking on a result item.
### 4.3.21 Top Entities service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Top Entities service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>The Top Entities service lists the places, people and organizations which occurred most often in the news in a certain time frame. The time frame and the starting day of the time period are defined by the user. In addition, a set of filters are available:</td>
</tr>
<tr>
<td></td>
<td>• The user can either ask for a top list in all news or specify a certain domain (like motorvision.de as shown in the screenshot below).</td>
</tr>
<tr>
<td></td>
<td>• All types or one of the specific types (Person, Place, and Organizations) can be displayed.</td>
</tr>
<tr>
<td></td>
<td>• The user can choose between German and English sources.</td>
</tr>
<tr>
<td></td>
<td>The Top Entities list is based on an in-depth analysis of millions of documents from more than 1.000 different news sources which goes far beyond a simple search for proper names using keywords.</td>
</tr>
<tr>
<td></td>
<td>The analysis module uses comprehensive lexica for the detection of &quot;candidates&quot; which might refer to a proper name. Lots of proper names are ambiguous. During document analysis, those names are disambiguated with the help of an analysis of the context – we use statistics generated from Wikipedia and a knowledge base derived from Freebase⁹ for this purpose.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC3</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>NEO</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://dicode-project.eu:34080/dicode-entity-prominence/pages/top-entities-widget.html">http://dicode-project.eu:34080/dicode-entity-prominence/pages/top-entities-widget.html</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>User interface description:</strong></td>
<td>The interface of the Top Entities service in the workbench is as follows:</td>
</tr>
</tbody>
</table>

---

⁹ [http://www.freebase.com](http://www.freebase.com)
Example of use:

Frank, Brand Manager for Mercedes-Benz, wants to know the top brands in the German motor press. He is especially interested in the magazines “Motorvision” and “Autobild”. First he has a look at the top organizations in both magazines for the entire year. Afterwards he looks at the results of the last months. Frank finds out that the position of Mercedes-Benz varied throughout the year. Therefore he asks Alice (Social Media Analyst) to analyze the relevant articles via the Entity Prominence service.

*Figure 1. User interface for the Top Entities service*
### 4.3.22 Top Hashtags on Twitter service (Keytrends service)

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Top Hashtags on Twitter service (Keytrends service)</th>
</tr>
</thead>
</table>
| Description: | Twitter users tag their tweets in a typical way: they use a term preceded by the hash character #. Those hashtags are often shared by many users. This service presents the top hashtags of a day as a rotating tag cloud. The service aims at helping the analyst to get a broad impression of the most important trends and topics on a certain day.

Most tag clouds show that hashtags come in different flavours. Many hashtags deal with topics of limited public interest, but most globally important events will pop up in the top hashtags list. Depending on the context, a hashtag can be “invented” collaboratively by a group of Twitter users. Organizers of public events often suggest an appropriate hashtag which is then adopted by the community. Some hashtags are used over a longer period of time. Some conventions concerning hashtags have been established over the years.

The top hashtags service is based on a representative sample of tweets from the Twitter streaming API. A blacklist is used for the removal of offensive terms. |
| Type of integration: | Light integration |
| Useful for: | UC3 |
| Developed by: | NEO |
| Additional information: | - |

#### User interface description:

The interface of the Top hashtags service in the workbench is as follows:

![Top hashtags from Twitter](image)

*Figure 1. Visualization of the top hashtags from twitter data*
By default the Top hashtags service shows the day before the current date. The user can adjust the date by using the date picker in the upper left corner.

![Top hashtags from Twitter](image)

*Figure 2. Filtering top hashtags by date*

**Example of use:**

Alice monitors social media for a large number of customers. At the start of each day, Alice wants to know what Twitter users are talking about, because she is convinced that tweets reflect current trends and that some Twitter users even are trend setters in social media. Alice has a look at the top hashtags. If she finds an interesting hashtag, she searches Twitter for all new tweets containing the hashtag.
4.3.23 Topic Detection service

Service name: Topic Detection service

Description: Topic detection provides graphical overview of the topics that are prevalent in a text collection

Type of integration: Light integration

Useful for: UC3

Developed by: FHG

Available at: http://tmserver.iais.fraunhofer.de:9090/TopicDetectionService

Additional information: -

User interface description:

The service computes a topic model based on the indicated text collection. The content of the topic model is coded as a topic graph – either in jpeg or in Gephi\textsuperscript{10} format. The graph can be downloaded by the user. The details are explained in the example of use.

The parameters of the topic detection service are explained in detail in the following section:

- **Text Collection Source**: The source of the texts on which the topic model is built. The ZIP archive must contain ASCII texts and may be sub-structures (see above).
- **Terms of interest file**: This file contains a list of terms which are of interest to the user. The terms may contain regular expressions (aka UNIX grep), and they are separated by “#”. This may be used to enrich the graph by meta-information. In our car review example, terms of interest might be “fuel consumption # speed # price # comfort”, etc. The terms of interest are displayed in the same fashion as the “text categories” (see above and figures 1 and 2).
- **Topic Model Parameters**: The topic model can be either small and with only more general topics (“coarse grained”) or large with detailed topics (“fine grained”).
- **Graph Format**: The graph display format (see above)
- **Specify Language**: The user must specify the language of the text collection in order to activate an appropriate filter that removes terms without information (“stop words”) before building the topic model.
- **Other options**: The user decides whether to display text categories and terms of interest (if appropriate). He or she may also restrict the number of topics to be displayed. Leaving empty a field of numbers means to display all items (which is of course eliminates the effect of the third option “Select top topics”).
- **Execution**: start the service.
- **Cancel**: cancel the service.

\textsuperscript{10} https://gephi.org
Example of use:

To activate the service, the user first has to specify the source of the text collection. He or she can either upload a ZIP archive or specify a date range in the text harvester to harvest either tweets or blogs. The ZIP archive must contain ascii text files. It can have a structure of sub-directories. If this is the case, the sub-directories are interpreted as a classification of texts by the system. The user then has the option to display the text categories in the graph (see “other options”).

Our example is a text collection of car reviews. The sub-directories are named after the cars being reviewed, and the car names appear in the graph display of figure 2. In this case the user indicated to display all categories (by leaving the field “select top topics” empty), and to only display the top 20 car names (by setting the field “number of text categories to display” to the appropriate number. Setting the field “select top topics” to 20 produces the graph in figure 3.

The resulting graph is either displayed as a jpeg picture (“Graph Format JPEG”), or downloaded as a Gephi file (“Graph Format Gephi”). The user can display and explore the Gephi graph locally using the free Gephi tool.

### Topic Detection

<table>
<thead>
<tr>
<th>Text Collection Source:</th>
<th>○ upload ZIP archive</th>
<th>○ Query text harvester</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time interval:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Start date:</td>
<td>01  2011</td>
</tr>
<tr>
<td></td>
<td>End date:</td>
<td>02  2011</td>
</tr>
<tr>
<td></td>
<td>collection type:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>○ Twitter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>○ Blogs</td>
<td></td>
</tr>
<tr>
<td>Terms of interest file:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic Model Parameters:</td>
<td>○ build a coarse-grained topic model</td>
<td>○ build a fine-grained topic model</td>
</tr>
<tr>
<td>Graph Format:</td>
<td>○ JPEG</td>
<td>○ save Gephi file (enter Path)</td>
</tr>
<tr>
<td>Specify language:</td>
<td>○ English</td>
<td>○ German</td>
</tr>
<tr>
<td>Other options:</td>
<td>□ Display text categories - number of categories to display: 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Display terms of interest - number of terms to display: 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Select top topics - number of topics to display: 10</td>
<td></td>
</tr>
<tr>
<td>Execution:</td>
<td>Go!</td>
<td></td>
</tr>
<tr>
<td>Cancel:</td>
<td>[X] Close Window</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1. Topic detection user interface*
Figure 2. Car review topics
Figure 3. 20 topics selected from the car review graph
5 Conclusions & Future Work

This deliverable is an enhanced version of deliverable 5.4.1, presenting new advances and developments carried out mainly in tasks 5.3 and 5.4 during the second year of the project. The new version of the Dicode workbench has been presented together with its novel features such as, for instance, user interface customization, expansion of widget to the center and elastic web design. More details about the generic integration framework have been provided considering three major topics: flexibility, scalability and sustainability. The two integration approaches defined for Dicode have been introduced: i) light integration, and ii) full integration. We defined two different approaches to facilitate the integration of most web applications, both applications developed within Dicode project as third parties tools.

As a central issue, the suite of services and applications integrated in the Dicode workbench has been presented according to a template also introduced in this deliverable. This deliverable includes descriptions of 23 services of which 6 implement full integration and 17 light integration.

The envisioned plan for the last year of the project includes:

- the complete definition of the integration framework, establishing the final technical features for integrating services
- the integration of more services from both inside and outside (third parties) of the project
- the evaluation of the workbench and the integrated services by end-users

In month 34 (end of July 2013), a final version (deliverable 5.4.3) of the current deliverable will be produced including all relevant aspects from the last period of the project.
References


