The Dicode project ([dicode-project.eu](http://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
Summary

This deliverable reports on the developments done within Tasks 5.3 (Dicode user interfaces) and 5.4 (Meaningful technical integration of Dicode services) of WP5. It constitutes the final version of a series of three deliverables reporting on the suite of services developed during the Dicode project and integrated into the workbench. This final version contains updates of services already presented in previous versions and new services developed during the third year of the project. Additionally, the new functionalities and features of the Dicode workbench are detailed. As in the previous versions, the intended audience of this document is end users of the Dicode environment. All service descriptions are given from a user-oriented point of view. Technical details of the presented services are not included, since they have been already reported in other final deliverables (D3.2.3, D4.1.3 and D4.2.3).
Table of Contents

1 Introduction .................................................................................................................. 5
  1.1 Context and 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
    2.2.1 List of recommended services ...................................................................... 6
    2.2.2 Browser compatibility chart .......................................................................... 7
    2.2.3 New workbench layouts ................................................................................. 7
    2.2.4 New widgets interface ................................................................................... 10
    2.2.5 New states for the widgets ............................................................................. 11
    2.2.6 Help and documentation of services .............................................................. 11
    2.2.7 New field for publishing services ................................................................. 12
    2.2.8 Unification of Titles and Forms .................................................................... 13
    2.2.9 Adding new services ....................................................................................... 13
    2.2.10 Tooltips ........................................................................................................ 14
    2.2.11 Video casts ................................................................................................... 14

3 Suite of integrated services .......................................................................................... 15
  3.1 Introduction ........................................................................................................... 15
  3.2 List of integrated services ..................................................................................... 15
    3.2.1 Collaboration service ...................................................................................... 16
    3.2.2 Crawl Visualization service .......................................................................... 22
    3.2.3 DBpedia Spotlight Named Entity service ..................................................... 24
    3.2.4 Emotion Detection Training service .............................................................. 28
    3.2.5 Entity Prominence service .............................................................................. 32
    3.2.6 PubMed Central Graph service ...................................................................... 34
    3.2.7 Recommender service for GEO datasets ...................................................... 40
    3.2.8 Similarity Learning service for GEO datasets .............................................. 42
    3.2.9 Subgroup Discovery service for functional interpretation of gene data ....... 45
    3.2.10 Top Entities service ...................................................................................... 47

4 Conclusions .................................................................................................................. 49

References ....................................................................................................................... 50

APPENDIX A: New REST calls for Collaboration services ............................................. 51
1 Introduction

1.1 Context and objectives

This deliverable constitutes the final version of a series of three deliverables reporting on the suite of services developed within the Dicode project. It is mainly focused on the work done in WP5 (Integrated Services & Innovative Work Methodologies), and specifically in Tasks 5.3 (Dicode user interfaces) and 5.4 (Meaningful technical integration of Dicode services).

The previous versions of this deliverable (D5.4.1 and D5.4.2) presented the outcome of development efforts carried out during the first 28 months of the project. This deliverable reports on the work done during the last six months. It is focused on two main issues: (i) the new functionalities and features of the Dicode workbench, and (ii) an updated list of services describing the new ones as well as the modifications done in previously reported services.

This deliverable follows the same approach used in the previous ones of this series, i.e. the descriptions of the workbench and services have been done considering the intended audience, namely the Dicode end-users. This is the reason why the service descriptions look like a user manual.

1.2 Structure of the deliverable

This deliverable follows the same structure as the previous ones. We have tried to avoid repetition of contents previously stated in the initial (D5.4.1) and the enhanced (D5.4.2) versions. A short introduction has been included in Sections 2 and 3 with two objectives: first, to summarize the contents stated in the related section of the previous versions and second, to introduce the new contents developed during the current reporting period.

Section 2 describes the new features and functionalities of the Dicode workbench. Section 3 contains the suite of services developed in the Dicode project. This is not an exhaustive list, since it reports only on the updates of existing services and new services developed during the last six months. Section 4 presents the final conclusions about integration of services in Dicode. Finally, there is a section devoted to references and one appendix reporting on technical issues not detailed before elsewhere.
2 Dicode Workbench

2.1 Introduction

The Dicode workbench is a widget-based web application that allows users working collaboratively by using Dicode services under a common interface. It constitutes the main product created within WP5. Introduced in D5.4.1, the Dicode workbench has evolved during the project to adapt its functionalities and aspects to the project’s evolving needs.

Changes in the Dicode workbench have been carried out in order to improve its usability and the user experience. These changes were motivated by the comments received from the project’s reviewers and the Project Officer during the second Review Meeting, as well as by the users participating in the second evaluation round (see D6.2.2, D6.3.2 and D6.4.2 for details). All the above reported that the usability of services should be enhanced to maximize the impact of the results. Most of the suggestions and recommendations raised by the reviewers and the users regarding the Dicode workbench have been addressed during the last period. This section reports on all the changes implemented in the Dicode workbench on this regard.

2.2 New features in the Dicode workbench

During the last period of the project, different maintenance tasks have been carried out, mainly fixing bugs and errors detected during testing. Apart from these issues, new features have been implemented and integrated to improve the usability of the Dicode workbench. The following subsections describe these changes in greater detail.

2.2.1 List of recommended services

The Dicode workbench enables users to register and publish services dynamically. Unfortunately, most users were unaware of the new registered and published services unless they were directly contacted by the publisher of the service. To alleviate this problem, a personalized list of recommended services has been implemented. Each user is presented with a different list of services. This list is displayed in the welcome page of the Dicode workbench after user logs in. Figure 2.1 shows how the list of recommended services is

![Figure 2.1 - List of recommended services.](image-url)
displayed to the user.

Figure 2.1 shows a selection of three services according to different factors such as, for instance, user preferences, popularity of services or publication date of the service. The algorithm that calculates the recommendations first considers the publication date of the services, giving more relevance to the more recent ones. Then, the algorithm considers the popularity of the service, i.e. how many times a service has been used in workspaces. The more a service is used, the more preferred it gets. Finally, the algorithm takes into account the annotations of services and gives more relevance to services that are similar to those already exploited by the user.

### 2.2.2 Browser compatibility chart

In web development, the compatibility of applications with different web browsers constitutes a great challenge because each browser works in a different way. With respect to the Dicode workbench, we have tried to maximize such compatibility by supporting five major browsers. Unfortunately, it has not been possible to achieve complete compatibility of all features with all of them. To inform users about this compatibility, a new compatibility chart has been included in the welcome page, below the list of recommended services, as presented in Figure 2.2. This list is the same for all users.

![Browser Compatibility Chart](image)

**Figure 2.2** – Web browser compatibility chart.

### 2.2.3 New workbench layouts

One of the main concerns reported by both service developers and end users was that the available space for displaying services was too confined. To provide a more intuitive and usable interface to end users, the layout of the workbench has been redesigned. Until now, the working space was logically distributed into three columns: two columns to display minimized services and a big central column for one maximized service.

Two new options for the layout were analyzed and discussed by the consortium: first, a two-column layout with one column aimed to minimized services and the other column for a maximized service (see Figure 2.3). The second option was a layout with a unique big central column where only one service is displayed at a time (see Figure 2.4). Finally, the Consortium decided to adopt both, enabling users to exchange between layouts dynamically.
To exchange between layouts, there are two clickable icons ( ) located in the upper right corner of the working space.
The main features of the two-column layout are (see Figure 2.5):

- 100% screen width use
- One left column containing minimized services and one central-right column for one maximized service
- Two icons at the upper right corner for enabling users to exchange between layouts
- Collapsible services in the left column for an efficient use of the space
- Toolbar on top

![Figure 2.5 – Workspace distribution in the two-column layout.](image)

This layout is aimed to users who usually use various services and need to exchange and share information between them.

The main features of the one-column layout are (see Figure 2.6):

- 100% screen width use
- Services are not presented as widgets but as a list on top of the page
- Toolbar at the bottom
- Maximum use of the screen space in the central area
This layout is oriented towards users who intend to work with services handling large amounts of data and need as much space as possible for visualization purposes; in this case, visual exchange of data (drag & drop items) between services is not needed.

2.2.4 New widgets interface

The appearance of widgets has been updated to provide users with more functions. This new interface is only valid in the two-column layout. The new appearance is shown in Figure 2.7.

Figure 2.6 – Workspace distribution in the one-column layout.

Figure 2.7 – New widgets interface for the two-column layout.
There are four buttons located on the upper-right corner of the widgets. Its functionality is:

- **Help button** – display the user help of the service (if available)
- **Exchange button** – maximize/move the service to the central area
- **Delete button** – remove the service from the working area
- **Expand/Collapse button** – expand or collapse the service (see below)

### 2.2.5 New states for the widgets

In the two-column layout, services located on the left column (Services Area) can be expanded and collapsed by users by clicking on the buttons presented in the previous subsection. When the widget is expanded, the service is presented as in the previous version of the Dicode workbench. When the widget is collapsed, the body of the service is hidden and the user can only see the title bar. Figure 2.8 illustrates both states.

![Collapsed State vs Expended State](image)

**Figure 2.8** – State of the widgets: collapsed – expanded.

### 2.2.6 Help and documentation of services

By clicking on the “Help button”, a floating window containing help information about the service is presented to the user. The information displayed on the floating window can be dynamically customized by the publisher of the service (see Subsection 2.2.7). If the publisher does not configure the user help, the system displays an error message indicating that there is not any help available. Figure 2.9 shows an example of the help pop-up window of a service.
2.2.7 New field for publishing services

A new field has been added to the publication form called “Help URL”. This field allows publishers to indicate an URL where the help of the service is available. This help should be in HTML format. It is presented when the user clicks on the “help button” of the service. Figure 2.10 shows the publication form with the new field.
2.2.8 Unification of Titles and Forms

To improve the friendliness and consistency of the interface, titles have been added to some forms and their visual style has been unified. Links providing help have been also added (see Figure 2.11).

![Figure 2.11 - Uniform titles and help links in Dicode workbench forms.](image)

2.2.9 Adding new services

The window for searching and adding available services has been also modified. The results are presented aligned. The link for adding services to the workspace is more visible in the right part of the dialog window (see Figure 2.12).

![Figure 2.12 - Floating window for searching and adding services.](image)
2.2.10 Tooltips

To further improve usability, tooltips have been added to most of the buttons in the Dicode workbench. Tooltips appear when users pass the mouse pointer over the buttons, presenting short help messages about the functionality of the item.

2.2.11 Video casts

During the second evaluation round, some video casts were produced to provide users with simple and attractive demonstrations on how to use the Dicode workbench and its possibilities. Links to some of these video casts have been included in the welcome page of the workbench.
3 Suite of integrated services

3.1 Introduction

This section reports only on the new services (developed during the last period of the project) and the updates done in previously presented services. For that reason, the list of services described in this section is not exhaustive, since we tried to avoid the repetition of contents. To describe services, we have used the template introduced in D5.4.2.

To simplify the understanding and facilitate the traceability of services, Table 3.1 summarizes the services reported in this section. For each service, it is indicated whether it is new or has been updated (during the last six months). Services that are not explicitly mentioned in this section remain unchanged (their description can be found in D5.4.2).

<table>
<thead>
<tr>
<th>Name of the service</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration service</td>
<td>Updated</td>
</tr>
<tr>
<td>Crawl Visualization service</td>
<td>New</td>
</tr>
<tr>
<td>DBpedia Spotlight Named Entity service</td>
<td>New</td>
</tr>
<tr>
<td>Emotion Detection Training service</td>
<td>Updated</td>
</tr>
<tr>
<td>Entity Prominence service</td>
<td>Updated</td>
</tr>
<tr>
<td>PubMed Central Graph service</td>
<td>New</td>
</tr>
<tr>
<td>Recommender service for GEO datasets</td>
<td>Updated</td>
</tr>
<tr>
<td>Similarity Learning service for GEO datasets</td>
<td>Updated</td>
</tr>
<tr>
<td>Subgroup Discovery service for functional interpretation of gene data</td>
<td>Updated</td>
</tr>
<tr>
<td>Top Entities service</td>
<td>Updated</td>
</tr>
</tbody>
</table>

Table 3.1 – Status of the services reported in the deliverable.

Apart from the services listed in this table, there is a set of four services (whose functionality remains almost the same as described in D5.4.2) that have now been integrated in (and an operational version of them is available through) the Dicode workbench. These services are:

- Opinion Mining service
- Phrase Extraction Application service
- Phrase Extraction Training service
- Topic Detection service

3.2 List of integrated services

In this section, the list of the services implemented or updated during the last period is presented. Services are listed in alphabetical order.
3.2.1 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 (see D4.1.1, D4.1.2 and D4.1.3). 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.

**Main menu**

The main menu of the Mind-map view has been relocated and improved (Figure 3.1). The Workbench user may access the menu by hovering on the left side of the collaboration workspace; a number of options are provided through the menu including access to the implemented multi-criteria algorithms, the workspace statistics, the creation of workspace objects, the creation of adornments, the replay mechanism, the workspace notifications, the filtering mechanism, and the proactive help and suggestions.
Figure 3.1 – Location and style of the new main menu in the final version of the collaboration service. The main menu is located on the far left side of the Mind-map view and individual options pop-out when users place the mouse over them.

Filtering a workspace

The final version of the collaboration service (see D4.1.3) introduced a new filtering feature allowing users to filter out items that appear on the collaboration workspaces. The filtering feature is accessible through the main menu of the collaboration workspace (in particular the menu button ). When pressing the menu button, a window pops up (Figure 3.2) to allow users define a number of criteria (such as the item’s title, its creator, the user who last modified the item, the date that the item was created or last modified, the mime and knowledge type of the item, its argumentation level etc.); pressing the button “Filter”, an updated instance of the workspace appears, containing the collaboration items that match the selected criteria.
Figure 3.2 – User interface dialog which allows users to select criteria to filter items in a collaboration workspace operated in the Mind-map view.

Proactive help and suggestions

A proactive help mechanism has been implemented to further assist users understand and exploit the capabilities of the Mind-map view of collaboration workspaces. The relative pop-up window (Figure 3.3) is accessible through the main menu of the collaboration workspace (by pressing the button); upon pressing this button, a window appears displaying a help message to provide assistance concerning a specific functionality of the collaboration workspace. The user may exploit the buttons at the bottom right side of the window to close it or display another help message.
Sharing of collaboration workspaces via social networking sites

Sharing of a collaboration workspace is possible through the “Share” menu (Figure 3.4) on the left bottom corner of the collaboration workspace. Clicking on the menu, a number of options appear corresponding to popular social networking sites (including Facebook, Twitter, LinkedIn, etc.). By clicking on a menu option, a workspace user is moved to the relevant API of the social networking site to share the URL of the collaboration workspace. Even in this way of sharing, users require the appropriate rights to access the collaboration workspaces. This means that users with no appropriate rights cannot access or view these shared workspaces.
Figure 3.4 – Quick social networking sharing option of collaboration workspaces via the main menu.

Displaying content

The workbench user may reveal an item’s content (Figure 3.5) by hovering over the item’s icon placed on the Mind-map view of the collaboration workspace. The tooltip that subsequently appears includes the content and the title of the item.
Figure 3.5 – Example of displaying an item’s content when the user places the mouse over the item.

Example of use:

A web application will provide the necessary user interface through which the previously mentioned operations can be executed by end users. In general, all Dicode use cases that require support for collaboration will be able to use the above operations. In particular, users of the Dicode use cases can use the operations provided by the workspace interface to create and configure new workspaces where the collaboration will take place. When they use collaboration workspaces in the Mind-map view and the workspace reaches a critical level in terms of the number of uploaded items, they may filter the workspace by keeping only the items they are interested in.
### 3.2.2 Crawl Visualization service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Crawl Visualization service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>The Crawl Visualization Service provides graph-based visualizations of web crawls. Those graphs have to be prepared by Dicode’s developers and uploaded via Dicode’s storage service.¹</td>
</tr>
</tbody>
</table>

Different types of graphs will be made available. In order to understand the content of a graph, a description has to be provided. Typical graphs might be presenting the following information:

- **Link structure of general web crawls**: Like the graph depicted below, this type of graph shows the link structure (incoming links) of a web crawl. Typically, only the domains with the most incoming links and the domains linking to them are shown.

- **Link structure of a subset of web crawls** (e.g. documents classified as weblogs): This graph type only shows domains of the same class. Additional filtering might be required to gain a legible result. The user might request a graph showing the top 1000 business sites.

- **Link structure of the ego network of a single domain**: This graph displays the incoming links to a single domain. The size of the nodes represents the number of links of the domain to the domain in question. The color of the nodes represents additional information about a node, e.g. a classification like “Travel”, “Motor” or “Sport”.

A set of general graphs will be initially available.

<table>
<thead>
<tr>
<th><strong>Type of integration:</strong></th>
<th>Light integration</th>
</tr>
</thead>
<tbody>
<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>Used via Dicode’s storage service</td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

#### User interface description:

The following figure shows an example graph which might be uploaded on the workbench. The graph shows the domains with the highest amount of incoming links in a general web crawl.

¹ Experiments showed that the fully automatic generation of crawl visualizations did not lead to good results. For a discussion of this issue, see D4.1.3.
Example of use:

Frank, a brand manager for Mercedes-Benz, wants to figure out which are the top sites linking to the Mercedes-Benz web site. He also wants to monitor the top web sites over time. Therefore he requests a monthly visualization of the 100 most important sites linking to the Mercedes-Benz web site.

He notices a significant shift related to major events and campaigns. Some of the top domains remain the same over time – although the total number of links changes depending on the season or on events like launches of new models or car fairs. If a new domain shows up in the top 100, Frank visits the web site and has a look at the links. He gains additional insights from comparing the results of the Crawl Visualization Service with the report from the tracking tools used by his department which inform him about the provenance of the traffic to the Mercedes-Benz web site.
3.2.3 DBpedia Spotlight Named Entity service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>DBpedia Spotlight Named Entity service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>The DBpedia Spotlight Named Entity service automatically annotates proper names (Named Entities) in text and links them to DBpedia resources.</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>
<tr>
<td>Available at:</td>
<td><a href="http://spotlight.dbpedia.org/demo">http://spotlight.dbpedia.org/demo</a></td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the DBpedia Spotlight Named Entity service in the workbench looks as follows:

![DBpedia Spotlight Interface](image-url)
By default, an English text about Berlin is shown in the text field in the center of the widget. If the user presses the “Annotate” button, the text is annotated with named entities. If no changes to the other controls are made, the default settings are used. All annotated entities will be underlined and linked to DBpedia resources as shown in the following figure.

Via the links to DBpedia, the user can look up additional information about the entity. The DBpedia resource page displays all available information from DBpedia.

The UI of the DBpedia Spotlight Named Entity service provides the following controls:

- **Language**: The annotation algorithm is highly language specific. Currently the user can choose between English, German, French, Italian, Dutch, Russian, Spanish, Portuguese and Hungarian. After a language is selected, a corresponding example text is shown.

- **Annotation score**: The annotation score parameter and the confidence parameter (described below) can be used to control precision and recall, i.e. the trade-off between error tolerance and completeness of the annotations. This trade-off is highly dependent upon specific use cases. Sometimes the correctness of the annotations is most important, while in other instances all findable concepts are of interest, even if this means more disambiguation errors.

Annotation score is a parameter that can be used to restrict shown entities to the ones that have a high “contextual” score. This means that the entity occurs in a familiar context (for example the singer Madonna in a text about her latest album and not in a text about cooking). When large values of annotation score are selected, the application does not annotate terms that have only little topical pertinence with respect to the given text.
• **Confidence:** The confidence parameter can be used to enforce a certain gap between the annotation score of the first and second ranked entity from the disambiguation process. The gap must be bigger than a relative threshold in order to be included in the result set. In this way, only annotations for which there is a clear single concept that can be chosen in a given context will be shown.

• **n-best candidates:** If checked, the application returns a ranked list of possible disambiguations instead of only returning the top ranked one. This option can be used to get an idea of the ambiguity of names and phrases and to understand the challenges in the annotation process.

• **Select types:** The “Select types” control allows for the selection of a custom set of entity types, as shown in the following figure:

![Select types](image)

**Example of use:**

Tina, a marketing professional, starts working on a market analysis about the music industry. She finds a comprehensive report about recent trends in global music development. The text reports on local, national and transnational genres emerging and
mentions a huge number of artists Tina is not familiar with.

She uses the DBpedia Spotlight Named Entity service to annotate the text with music related metadata which she selects via the “Select types” menu. The service returns links to DBpedia resources describing genres and artists. Tina quickly checks the results and informs herself about the newest trends and the emerging stars of the music industry including interesting facts and relations about them.
3.2.4 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. Optionally a list of unwanted terms can also be included (with regular expressions) for post processing the detected 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>Dicode workbench</td>
</tr>
<tr>
<td>Additional information:</td>
<td>-</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). Optionally a second list of unwanted terms can be supplied. An exhaustive example is given below. Please note that the lists may be edited and re-used to train an even better model with the Phrase Extraction Training service.

The parameters of the Emotion Detection Training service are explained in detail here:

- **Upload text collection (ZIP archive):** 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 wanted terms file (regular expressions):** This list is an enumeration of so-called seed terms. These should mark phrases that contain the emotion that the user wants to be extracted (for example “I like” in English). The terms may be single words or short phrases (idiomatic expressions). They may contain regular expressions. Each term is enclosed in double quotes and the terms are separated by commas.

- **Upload unwanted terms file (regular expressions):** This list is second enumeration of seed terms. These should mark phrases that do not contain the emotion that the user wants to be extracted, but would match the positive seed terms (for example “I don’t like” in English). The terms may be single words or short phrases (idiomatic expressions). They may contain regular expressions. Each term is enclosed in double quotes and the terms are separated by commas.
- **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 field “Show extracted phrases after completion” will prepare for download the text collection with highlighted phrases (considering the size of the training collection, this will be very lengthy!). Tagging the field “Show tag cloud after completion” will prepare for download a tag cloud that is built based on all the detected phrases.

- **Execution**: start the service.

- **Cancel**: cancel the service.

---

**Emotion Detection Training**

<table>
<thead>
<tr>
<th>Upload text collection (ZIP archive)</th>
<th>Browse...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upload wanted terms file (regular expressions)</td>
<td>Browse...</td>
</tr>
<tr>
<td>Upload unwanted terms file (regular expressions)</td>
<td>Browse...</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, German</td>
</tr>
<tr>
<td>Other options:</td>
<td>Show tag cloud after completion, Show extracted phrases after completion</td>
</tr>
<tr>
<td>Execution:</td>
<td>Go!</td>
</tr>
<tr>
<td>Cancel:</td>
<td>[X] Close Window</td>
</tr>
</tbody>
</table>
Example of use:

A user wants to train the system to detect phrases that express an emotion of content or joy in German texts in the automotive domain. She uploads a ZIP archive of 2000 postings from the German Web forum “Motortalk”.

Second she uploads a file that contains a list of the wanted seed terms. Here we show an excerpt from the file to demonstrate the character of the seed terms:

```
"([Ff]|([Pp]h)antastisch",
"(^|
 )([Ww]under|([Aa]ller|[Bb]ild)*schön(e|es|ne|em|er|eres|eren|erem|stes|stem|sten|stem)($| )",
"(^| )Cool(st)*e|en|er|em|es)*($| )",
"(^| )Schön(e|es|ne|em|er|ere|eres|eren|erem|ste|stes|ster|sten|stem)($| )",
"(^| )Spa(ss|ß)($| )",
"(^| )[Ff]reu(e|st|en)( |$",
"(^| )[Gg]lücklich($| )",
"(^| )[Ww]under (bar|voll)(er)*(e|es|er|en|em)*($| )",
"(^| )gefallen($| )",
"(^| )gefällt($| )",
"(^| )klasse($| )",
"(^| )prima($| )",
"(^| )super($| )",
"(^| )toll(e|es|er|en|em)*($| )",
"(ober|super|mords| | )cool(st)*e|en|er|em|es)*($| )",
"[Bb]egeistert",
"[Ee]mpfehlenswert",
"[Gg]eil",
"[Hh]euerlich",
"[Hh]ervorragend",
"[Ss]ensationell",
"[Zz]ufrieden",
"hervorragend"
```

The seed terms are enclosed by double quotes and separated by commas. The regular expressions subsume declination and conjugation alternatives of German words as well as alternative modifiers. The expression "(^| )super($| )" notes that the term “super” should occur either at the start or end of a phrase, or it should be enclosed by whitespace.

Then, the user adds a third file of expressions that match the wanted terms, but should nevertheless be excluded. These expressions include negations (“nie”, “nicht”) and other modifiers that mitigate the positive emotion (“zu schön”, “zwar”). We also give an example here:

```
"(zu|m)al) schön ",
" net ",
" net$",
" nicht",
" nie",
" so schön",
" zwar",
"([Aa]uf|[Üü]ber
```
She gets an extraction model and a list of phrases detected by the extraction model. She uses the phrases that are inappropriate to develop modifications of the regular expressions of both the wanted terms and the unwanted ones and starts a second Emotion Detection Training process.

Alternatively, she could continue with the Phrase Extraction Training and the Application services. First she uploads the extraction model and another collection of 1000 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 3000 posting texts to the Phrase Extraction Training service to train a new and better model.
### 3.2.5 Entity Prominence service

<table>
<thead>
<tr>
<th>Service name</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.</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>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

#### User interface description:

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

![Entity Prominence service interface](image)

The interface uses colors to improve the usability. On the top of the left column, the user sees the selected entities and the parameters (like “François Hollande” and “Source: Tagesspiegel”). The same colors are used for the border of the respective prominence graph. The background of the graph is filled transparently with the same color. An
additional legend at the bottom helps the user understanding the graph. It also comes handy if the user wants to use the graph externally, like in a presentation.²

<table>
<thead>
<tr>
<th>Example of use:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remains unchanged (see D5.4.2).</td>
</tr>
</tbody>
</table>

² For presentations, an export functionality concerning both visualizations and data would be convenient. If resources are available, we will consider implementing this feature.
3.2.6 PubMed Central Graph service

<table>
<thead>
<tr>
<th>Service name:</th>
<th>PubMed Central Graph (PMC Graph) service</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>This service mines the content of the articles included in PubMed Central. Articles have been indexed considering the appearance or mention of terms belonging to controlled vocabularies. Currently, the vocabularies used are the following:</td>
</tr>
<tr>
<td></td>
<td>• Gene Ontology(^3) (GO) genes</td>
</tr>
<tr>
<td></td>
<td>• HUGO(^4) genes</td>
</tr>
<tr>
<td></td>
<td>• MeSH(^5) diseases</td>
</tr>
<tr>
<td></td>
<td>• MeSH tissues</td>
</tr>
<tr>
<td></td>
<td>• UniProt(^6) proteins</td>
</tr>
<tr>
<td></td>
<td>• KEGG(^7) drugs</td>
</tr>
<tr>
<td></td>
<td>• KEGG disease</td>
</tr>
<tr>
<td></td>
<td>The service allow users to perform queries with the following parameters:</td>
</tr>
<tr>
<td></td>
<td><strong>INPUT Terms:</strong> Gene Ontology genes &amp; HUGO genes</td>
</tr>
<tr>
<td></td>
<td><strong>OUTPUT Term(s):</strong> MeSH Diseases</td>
</tr>
<tr>
<td></td>
<td>Selecting one or more genes as input, and choosing one category as output, the service search for PubMed Central publications where those terms appears simultaneously. The results are presented as a graph.</td>
</tr>
<tr>
<td><strong>Type of integration:</strong></td>
<td>Light integration</td>
</tr>
<tr>
<td><strong>Useful for:</strong></td>
<td>UC1, UC2</td>
</tr>
<tr>
<td><strong>Developed by:</strong></td>
<td>UPM</td>
</tr>
<tr>
<td><strong>Available at:</strong></td>
<td><a href="http://hodgkin.dia.fi.upm.es:8080/pmcgraph/">http://hodgkin.dia.fi.upm.es:8080/pmcgraph/</a></td>
</tr>
<tr>
<td><strong>Additional information:</strong></td>
<td>-</td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the PMC Graph presents two slightly different layouts. One in case the service is minimized or presented as a widget and another one in case the service is on central area of the Dicode workbench. Following, the two layouts are presented together with a description of how the results are presented to users.


\(^4\) Human Genome Organisation (HUGO) – [http://www.hugo-international.org](http://www.hugo-international.org) ; [http://www.genenames.org](http://www.genenames.org)


\(^6\) Universal Protein Resource (UniProt) – [http://www.uniprot.org](http://www.uniprot.org)

Minimized layout

This layout is pretty simple as shown in the next figure:

Users have to specify the list of genes they are interested in. The interface provides two different methods to do that by clicking on the top buttons:

- **Input Genes** – selected as default, this option allows users to type manually the name(s) of the genes or the GO identifier. To facilitate user’s work, this autocomplete field provides users with different options while typing as presented below.

- **Load File** – when selected, a prompt box appears enabling users to upload a file containing the list of genes. The file must be a .TXT file with a list of gene names separated by commas.
Below, three more controls appear in the interface of the service:

- **Gene list box** – this box will be automatically populated with the genes established by users. Once in the box, users can remove them by clicking on the “x” button shown below.

- **Output table** – users are presented with a combo box to select the terms that they want to obtain relations with.

- **Search & Reset data buttons** – the search button executes the query, and the reset data button cleans the form to start a new query. It is important to note that when service is displayed using this layout, results are not shown because of the reduced space. User is notified about this and invited to maximize the service to see the results.

**Maximized layout**

This layout is almost the same as the previous one. It is mainly oriented to display the results of the queries. It presents two additional controls:

- **Export to .csv file** – this button allows users exporting the complete results of a query to a standard Comma-Separated Values (CSV) file. By default it is disabled but it is enabled after users execute a query with results.

- **Not related genes box** – this is an informative box. After users execute a query, this box is populated with the genes specified by users that have not got any relation with other terms and concepts.
Query results

Results are presented to users as a graph. The following picture shows an example of the result of a query.

In this graph, blue nodes represent genes identified with their GO Id and red nodes denote diseases. Arcs joining nodes represent relationships found between the terms represented in such nodes. Those relationships mean that there are publications, at least one, in which both terms are mentioned. Depending on the number of publications found, arcs are displayed using different colors as detailed in the legend.

Clicking on the arcs, the relationship info box is presented. This box contains information about the relation: the two terms related and the number of publications found mentioning both terms. To see the complete list of publications, there is a link “View Detailed Info” that, when clicked, a new floating window is opened presenting the publications (as shown in the following figure).

This window presents a table with the titles of the publications, publication dates and journal in which articles were published. Users are allowed to sort articles by any of these fields. Finally, titles of publications are links to the article in PubMed Central.
**Example of use:**

John is conducting an experiment in which he is analyzing the relationships between a couple of genes and general diseases. The name of the genes are GO:0000001, GO:0000004, and GO:0000006. He begins typing the name of these genes in the “List of Genes” text box. Thanks to the auto-complete feature, the user is shown the most probable genes, so he doesn’t have to write the complete name but select them from a list of possible options.

Once the genes have been added, John selects the type of terms with which the genes will be related to. In his case, he will select “MeSH diseases”

John executes the query clicking on “Search” button and he will be presented with the following graph as a result.

![Graph showing relationships between genes and MeSH diseases](image)

From the graph, and by passing the mouse for the various nodes, John finds interesting that the gene GO:0000004 is related with the node entitled “Death”. Thus, he clicks on the relation, and is presented with four articles that contain both of these terms. John wants to read the full article “Combining evidence, biomedical literature and statistical dependence: new insights for functional annotation of gene sets”, so he clicks on the title and he is referred to the PubMed Central page of the publication.
4. Discover an interesting relationship

5. Explore publications that relate two selected nodes

6. Get referred to the original page of the PubMed article
### 3.2.7 Recommender service for GEO datasets

<table>
<thead>
<tr>
<th><strong>Service name:</strong></th>
<th>Recommender service for GEO datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></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><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/SIIService/Recommendation.jsp">http://tmserver.iais.fraunhofer.de:8080/SIIService/Recommendation.jsp</a></td>
</tr>
</tbody>
</table>

**User interface description:**

The interface of the Recommendation of GEO datasets is presented below:

![Recommender service for GEO datasets interface](image)

The interface of the Recommendation of GEO datasets is presented below:
The user can formulate his needs in two different ways:

1. Select a GEO dataset of interest. The service delivers that datasets that are similar to the defined one. He has to specify the datasets by its unique GEO accession number.

2. Describe a dataset of interest in the corresponding text 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.

To start the recommender process, a model-ID is required. Model-ID is a unique number that identifies the similarity model which is learned in similarity learning service.

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.

**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.
3.2.8 Similarity Learning service for GEO datasets

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Similarity Learning service for GEO datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>This service aims to create a similarity model for comparison of GEO datasets</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>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. The similarity learning service starts by clicking on the “Train” link in the following interface:

![List of Similarity Learning Services](#)

The user is then forwarded to the learn process where he will be shown iteratively 15 pairs of datasets. For each dataset pair, the service displays side by side their representation in GEO repository.
For each dataset pair, the user is asked to mark them as “similar”/“dissimilar”/“don’t know”.

The service remembers the user feedback and learns a similarity model reflecting user preferences. After the model is successfully created the user is informed about this. Since it is allowed to create several models for different tasks, the user is asked to remember the model-ID in order to use this number later in recommender service.
**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.
3.2.9 Subgroup Discovery service for functional interpretation of gene data

<table>
<thead>
<tr>
<th>Service name:</th>
<th>Subgroup Discovery service for functional interpretation of gene 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 this service has been updated from the previous versión. Now it looks as follows:

![Subgroup Discovery on Gene Data interface](image)

Regarding fields in this interface, the difference with the previous version is that the three available ontologies are used by default. The rest of the information about this service remains unchanged.
### Example of use:

Remain unchanged (see D5.4.2).
### 3.2.10 Top Entities service

<table>
<thead>
<tr>
<th><strong>Service name:</strong></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. For a detailed description of the usage of the Top Entities service and of the Entity Prominence service see D5.4.2.</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>
</tbody>
</table>

#### User interface description:

A new design has been developed which tackles several usability issues of the former design. The functionality of the service remains unchanged. In the new design, a bar chart shows the amount of mentions for each entity. The usage of colors and legends makes it easy to understand the result on first sight.

Due to a shortage in front-end developer resources at NEO, the new design for the Top Entities service and the Entity Prominence has not been implemented yet.
Example of use:

Remain unchanged (see D5.4.2).
4 Conclusions

This deliverable constitutes the final version of a series of three deliverables reporting on the integrated services created during the Dicode project. It reports on the new services developed during the last six months of the project, how existing services have been updated and how the Dicode services were integrated into the workbench.

In total, more than 25 services have been developed by technical partners during the project to provide with innovative solutions to the problems faced by the use case (end-users) partners. All services have been integrated under a common web framework called the Dicode workbench. The Dicode workbench provides uniform access to all available services and enables users to work collaboratively. Mechanisms to dynamically integrate services in the Dicode workbench have also been implemented. The way chosen to integrate the Dicode services with the workbench enables also their stand-alone and independent use as traditional web applications.

To support the integration of services, two different approaches were defined: light integration and full integration. Light integration is aimed towards those services that do not need to exchange information with other services, i.e. work as independent entities. Full integration was defined to provide a common framework to allow services working together by exchanging information between them. More details about these two integration approaches appear in deliverable D5.4.2. Some Dicode services have already implemented both approaches; however, most of them are integrated using light integration. There are two main reasons for that: first, light integration requires fewer efforts from service developers, and second, most of the services required by end users do not need to interact and interoperate with other Dicode services. For the majority of end users, it is enough to be able to use different services within a common framework.
References


[D5.4.1] Dicode deliverable D5.4.1: Integrated Dicode Services (initial version).

[D5.4.2] Dicode deliverable D5.4.2: Integrated Dicode Services (enhanced version).

[D6.2.2] Dicode deliverable D6.2.2: Report from the evaluation of use case #1 (second version).

[D6.3.2] Dicode deliverable D6.3.2: Report from the evaluation of use case #2 (second version).

[D6.4.2] Dicode deliverable D6.4.2: Report from the evaluation of use case #3 (second version).
APPENDIX A: New REST calls for Collaboration services

In this appendix, we present some technical details regarding to the new REST services that have been implemented to improve the Collaboration service. These new REST services are listed in the following table. Although this deliverable is oriented to end users, we decided to include this technical information as an appendix of this deliverable since it has not been reported elsewhere before.

<table>
<thead>
<tr>
<th>REST SERVICE</th>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeleteUser</td>
<td>user=ken tmpstamp=20111229174139</td>
<td>Provides a user the option to delete the account of another already created workbench user (functionality provided only to Dicode workbench administrators). Functionality used for Dicode workbench user management.</td>
</tr>
<tr>
<td>UpdateWorkspace</td>
<td>id=25620 title=wrk_title descr=wrk_desc private=1 modify=1 upload=1 uploadmaxsize=1 type=22065 tmpstamp=20111229175536</td>
<td>Enables a user update the attributes of an already created collaboration workspace (including its title, its description, whether the workspace is private or not, whether file uploading is allowed, the maximum size of a file to be uploaded etc.). Rest call used when updating the attributes of a workspace is needed.</td>
</tr>
<tr>
<td>GrantUserAccess</td>
<td>user=wrk_user type=22065 tmpstamp=20111229175705</td>
<td>Enables a workbench user (having the appropriate rights) to grant access to another workbench user on a specific workspace.</td>
</tr>
<tr>
<td>RemoveUserAccess</td>
<td>user=testuser type=22065 tmpstamp=20111229175705</td>
<td>Used for removing access (that has been previously granted) from a user on a specific workspace.</td>
</tr>
<tr>
<td>getSessionUser</td>
<td>tmpstamp=20111229175705</td>
<td>Enables the caller to monitor who user (if any) is currently logged in the workbench under the same session. Functionality used to check whether any user is logged in the workbench under the same session (if not, the Login REST call should be called to Login the user in the workbench)</td>
</tr>
</tbody>
</table>

Table A.1 – Additional REST calls passed from the Dicode workbench to the Collaboration services.