Development and evaluation plan

D1.3

May 29, 2001, UNIDO/WP1/Version 3

Henrik Nottelmann, Alan Munro, Fabio Crestani
Contents

1 Development 2
   1.1 Framework ............................................................... 2
   1.2 Resource Description/Resource gathering .......................... 2
   1.3 Query transformation ............................................... 3
   1.4 Query modification .................................................. 3
   1.5 Resource Selection .................................................. 4
   1.6 Database query run/Data fusion ................................... 4

2 Evaluation 6
   2.1 Objectives ............................................................. 6
   2.2 User groups involvement ............................................. 7
   2.3 Usage scenarios ....................................................... 7
      2.3.1 Scenario 1: Website design for arts festival ............... 7
      2.3.2 Scenario 2: Newspaper feature ............................... 8
   2.4 Evaluation plan ....................................................... 8
Chapter 1

Development

This chapter describes the responsibilities (which partner is responsible for which class and method) and the date of delivery.

This document only covers the database-independent classes. In practice, three classes have to be subclassed to perform database-specific work (DLProxy, DLInterface and DLMediaComponent). These subclasses have to be implemented in a later phase of the project.

The responsible partners should deliver the specified code by the given dates in a usable version. Of course, further improvements (e.g., due to future research) are possible lateron.

If needed, dummy classes can be implemented before the dates of delivery to enable testing of components.

1.1 Framework

- Classes framework (all attributes are existent, all methods are existent but empty (returning only dummy values) – except getXYZ() and setXYZ() methods used for data hiding) ⇒ by end of June 2001

- communication “user interface – Dispatcher” and “Dispatcher – DLProxy” based on XML and TCP/IP sockets ⇒ by end of July 2001

- AbstractQuery, AbstractQueryCondition, Query, QueryCondition, PropQuery, PropQueryCondition, CostParameters, GlobalInformation ⇒ by end of June 2001


1.2 Resource Description/Resource gathering

- ResourceDescription ⇒ by end of August 2001

- DLProxy.gather() Task: applying query based sampling ⇒ by end of August 2001
• DLMediaComponent.extractResourceDescription(Result)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of August 2001

• extractResourceDescription(Result) in media-specific subclasses of DLMediaComponent
  ⇒ by end of November 2001

1.3 Query transformation

• DLProxy.getPropQuery(Query)
  ⇒ by end of July 2002

• DLMediaComponent.transformQuery(Query)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of July 2002

• transformQuery(Query) in media-specific subclasses of DLMediaComponent
  Task: extracting values for the resource descriptions and storing them in the resource descriptions w.r.t. their media type
  ⇒ by end of December 2002

1.4 Query modification

• Dispatcher.modifyQuery(Query)
  ⇒ by end of August 2001

• DLProxy.calcQueryModificationFeatures(Query)
  ⇒ by end of August 2001

• DLMediaComponent.calcQueryModificationFeatures(Query)
  ⇒ by end of August 2001

• DLMediaComponent.calcGlobalInformationFeatures(Query)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of August 2001

• calcGlobalInformationFeatures(Query) in media-specific subclasses of DLMediaComponent
  Task: generating global information features w.r.t. their media type
  ⇒ by end of November 2001

• DLMediaComponent.calcRelevanceFeedbackFeatures(Query)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of August 2001

• calcRelevanceFeedbackFeatures(Query) in media-specific subclasses of DLMediaComponent
  Task: generating relevance feedback features w.r.t. their media type
  ⇒ by end of November 2001

• DispatcherMediaComponent.modifyQuery(Query,QueryModificationFeatures)
  ⇒ by end of August 2001

• DispatcherMediaComponent.calcGlobalInformation(Query,Features)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of August 2001
• calcGlobalInformation(Query, Features)
  Task: generating global information values w. r. t. their media type
  ⇒ by end of November 2001

• DispatcherMediaComponent.applyRelevanceFeedback(Query, Features)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of August 2001

• applyRelevanceFeedback(Query, Features) in media-specific subclasses of DLMediaComponent
  Task: modifying the query w. r. t. the given relevance feedback data and their media type
  ⇒ by end of November 2001

1.5 Resource Selection

• Dispatcher.calcResourceSelection(Query)
  Main task: calculating an optimum selection
  ⇒ by end of December 2001

• DLProxy.calcExpectedCosts(Query)
  ⇒ by end of December 2001

• DLProxy.calcExpectedRelevantDocNum(PropQuery)
  Main task: estimating the number of relevant documents w. r. t. features
  ⇒ by end of December 2001

• DLProxy.calcExpectedPrecision(int, int, int)
  Main task: estimating expected precision
  ⇒ by end of December 2001

• DLProxy.calcExpectedCosts(int, double, CostParameters)
  Main task: estimating expected costs
  ⇒ by end of December 2001

• DLMediaComponent.calcRelevanceEstimationFeatures(PropQuery)
  Task: providing unique method for media-specific subclasses
  ⇒ by end of October 2001

• calcRelevanceEstimationFeatures(PropQuery) in media-specific subclasses of DLMediaComponent
  Task: generating features w. r. t. their media type
  ⇒ by end of April 2002

1.6 Database query run/Data fusion

• Dispatcher.queryDatabasesAndFuseResults(Query, int[])
  Main tasks: detecting duplicates, ranking the documents
  ⇒ by end of February 2002

• Dispatcher.retrieveDocument(Document)
  Task: retrieve complete document w. r. t. the given document
  ⇒ by end of July 2001

• Dispatcher.retrieveDocument(Result)
  Task: retrieve complete documents w. r. t. the given result
  ⇒ by end of July 2001
• DLProxy.query(Query, int)  
  ⇒ by end of July 2001

• DLProxy.retrieveDocument(Document)  
  Task: retrieve summary or complete document w.r.t. the given document  
  ⇒ by end of July 2001

• DLProxy.retrieveDocument(DocumentID)  
  Task: retrieve summary or complete document w.r.t. the given document id  
  ⇒ by end of July 2001

• DLInterface  
  ⇒ by end of July 2001

• DLProxy.queryDatabaseAndPostprocess(Query, int)  
  ⇒ by end of July 2002

• DLMediaComponent.postprocessResult(PropQuery, Result)  
  ⇒ by end of October 2002

• DLMediaComponent.recalcWeights(PropQuery, Result)  
  Task: providing unique method for media-specific subclasses  
  ⇒ by end of October 2001

• recalcWeights(PropQuery, Result) in media-specific subclasses of DLMediaComponent  
  Task: recalculating document weights w.r.t. their media type, filtering  
  ⇒ by end of February 2002

• DLMediaComponent.createSummary(PropQuery, Result)  
  Task: providing unique method for media-specific subclasses  
  ⇒ by end of October 2001

• createSummary(PropQuery, Result) in media-specific subclasses of DLMediaComponent  
  Task: creating a summary for attributes of their media type  
  ⇒ by end of February 2002

• DispatcherMediaComponent.getDuplicateCandidates(Result[])  
  Task: providing unique method for media-specific subclasses  
  ⇒ by end of October 2001

• getDuplicateCandidates(Result[]) in media-specific subclasses of DispatcherMediaComponent  
  Task: detecting duplicate candidates w.r.t. their media type  
  ⇒ by end of February 2002
Chapter 2

Evaluation

As indicated in the Technical Annex, each component of the MIND system will be thoroughly tested by each partner during the course of its development. When these components will have been integrated into the MIND system prototype, an overall system evaluation is needed to show how the developments of the MIND project help users in accessing distributed, heterogeneous multimedia digital libraries. This evaluation will draw on USG’s expertise in the field of user-centred task oriented evaluation and will take place within a large realistic setting involving a number of heterogeneous multimedia digital libraries.

The evaluation will draw on elements of the methodology developed by USG in the context of its participation to the ESPRIT working group Mira on the evaluation of interactive multimedia information retrieval applications. The final evaluation will be carried out in collaboration with the user groups identified in WP1 according to the evaluation plan delineated below.

It is important to notice that the full details of the evaluation plan cannot be known, at this stage, since there still are many open question regarding the scope/range of involvement of user groups at the different stages of the project, and the provision of utilising user feedback effectively and systematically. In addition, questions related to where, when and how often resource descriptions will be gathered and verified cannot be answered yet. All these details will be decided at a later stage and will be described in a later version of this document.

2.1 Objectives

The evaluation of the Mind system has a number of objectives. The two most important ones are:

- MIND evaluation will be part of an iterative design process which will inform the design of the system in tune with requirements of user-centredness, and usability.
- MIND evaluation will give a view on the capability of the system to fulfil the objectives of the MIND project.

In deliverable D1.2 we indicated a number of potential user groups. Agreements detailing the involvements of these user groups are still to be drawn. It should be noticed that other user groups may be added to those already indicated. A later version of this document will indicate other user groups for evaluation of the system. These user groups will be chosen to give a number of heterogeneous groups, united in their need for the kind of information which MIND is designed to deliver, but different enough in their particular circumstances to give a fair test of the versatility of the prototype system.

The selection of user groups will be related to the availability of digital libraries which MIND can use, and so the suitability of the information for true tests of a given user group’s needs. An
initial indication of these digital libraries is reported in deliverable D1.2. Further, selection will be driven by the generation of a number of user scenarios. User-scenarios have proved a useful way of generating design ideas, and also clarifying possibilities of system design and requirements. They are also useful in communicating from the design process to the various communities of practice, from designers and anthropologists down to coders and engineers.

2.2 User groups involvement

Mind will employ a user-centred and task oriented approach to the design process of the system throughout the project. It will do this by utilising a number of different methods from system design, such as HCI methods and methods from Participatory Design. We will target a number of different user groups, who will use the system in different ways and for different tasks. Partly, the user groups will be informed by the development of Design Scenarios. The scenarios will show examples of instances of use of the system, and will be rich enough to point to issues for design with the system, and suggest possible design opportunities. As said above, the requirements for user communities are twofold: to be different enough from each other to form a reasonable test of the system, but to be realistic: that is, to be users who might well be expected to use a system like MIND at least a few times a week, if not a day.

2.3 Usage scenarios

A number of possible use scenarios will be set out at a later stage. Some examples are given below. These example scenarios give a number of specific instances in which the MIND system will be expected to be used. They will be evaluated and iteratively worked on through the lifetime of the project, and in response to contact with user groups. Scenarios are numbered for easy reference to the various circumstances and instances which they refer to.

2.3.1 Scenario 1: Website design for arts festival

1. Patrick is writing a web page for a festival on Samuel Beckett, for an arts festival which will incorporate stagings of his plays, talks about him and his impact, showings of films about him, films of his plays, as well as of some select contemporaries. There will also be introductory exhibitions and events staged throughout the city.

2. Patrick has already got a lot of information about his life and oeuvre from biographical work and critical pieces on his plays. He is intending to use this as the main source for information on Beckett’s life, and position in 20th century literature.

3. However, he wishes to look at other materials. He needs to find what relevant materials are available to make the web site interesting and interactive. Also, what is not incorporated may be referred to and linked to.

4. He would like to find what is available. This might include pictures such as the famous series by Minihan, or any audio-visual material of interviews perhaps with Beckett, or with Beckett experts, such as Richard Ellman, or sound archives of his plays, interviews with him. He is also interested in any artwork by contemporaries which featured him.

5. The website design and festival itself are supported by the British Council and an action of the European Commission, as well as a number of other reputable sponsors, so it is important that the site does not contain any materials which might involve any copyright problems.

6. He goes to the MIND system and looks through what is available.
2.3.2 Scenario 2: Newspaper feature

1. Sarah is an editor for the “current events” section of a Sunday newspaper. This incorporates analysis, contextual and comment pieces of the week’s events, aiming to provide pieces on the week’s events as well as groups of pieces which are grouped around a particular news event.

2. She is trying to get a number of different pieces in response to a high-profile news story of the week, the Afghan Taliban announcing its intention to destroy a number of Buddhist monuments which are within Afghan borders.

3. She has a number of articles already written, but needs at least another one article, perhaps two on different issues related to this news event. She has a number of possibilities in mind after talking to colleagues who are more expert in Afghanistan.

4. There is the possibility of an article on the Taliban and their interpretation of the Koran, the relationship of Afghanistan to the west after some attempts at diplomacy by their government and in mind of action by the US, or an article on the Buddhists who built the monument, perhaps focusing on the territories throughout that region which were once Buddhist.

5. She can be flexible in terms of what is incorporated: It can either be a higher word count, or if she can find an arresting picture somewhere, a lower word count and a series of pictures from different sources. If there are a number of pictures which are especially good and they fit, she may get the word count of the articles edited down.

6. She wants to know what is available. Are there usable pictures which they might use? Are there articles or other information which she might incorporate in some way, either by using a syndicated article from another paper, or news agency? Is there any good visual material from picture agencies or stills which she might take from video material?

7. She uses the MIND system to find this out.

2.4 Evaluation plan

Evaluation will be carried out as task 6.3. This is proceeded by task 6.1 (integration) and task 6.2 (gathering of data for evaluation).

In task 6.1, the tools for resource selection and data fusion will be integrated into the MIND system prototype. In a preliminary integration, tools which cope with homogeneous resources only will be used. Later, when the components for dealing with heterogeneous resources become available, they will replace the prior ones. Task 6.1 is scheduled to start at PM 16 (April 2002) and end at PM 26 (end of February 2003).

In task 6.2 we will gather the different Digital Libraries needed to carry out the evaluation. The provision of data in different media will be responsibility of the partners that is mainly responsible for the specific media. So, DSI will construct and provide colour, texture and shape test image collections. USFD will provide audio and speech recognised collections. CMU will provide a set of tasks, comprising information needs, queries, interaction models, and a framework for the evaluation of the user satisfaction of such tasks. USG will provide the general framework for the task evaluation. UNIDO and DSI will be involved in defining the evaluation criteria: DSI looking at the construction and definition of suitable measures of retrieval efficiency and effectiveness in image retrieval; and UNIDO examining classical (batch-oriented) retrieval measures as well as modern user-oriented measures for consideration. Task 6.2 is scheduled to start at PM 25 (January 2003) and end at PM 26 (end of February 2003).

The actual evaluation will be carried out in task 6.3. As already pointed out, the evaluation will be user-centred and will be carried out under the direction of USG. Once the testing environment is in place and some suitable measures of performance are identified, USG will coordinate the user-centred evaluation of the prototype systems developed. USG will also co-ordinate the analysis of
the data gathered during the evaluation. Statistical analysis techniques will be used to isolate the
different variables involved in the evaluation and to test the effectiveness hypotheses formulated.
Such analysis will provide useful suggestions on how to improve the architecture, the interfaces
and the user-interaction styles of the prototype systems.

Three deliverables are involved in the evaluation workpackage (WP 6): D6.1 on the evaluation
and integration of the MIND prototype components, D6.2 on the design of the testbed for MIND,
and D6.3 reporting on the findings of the user-centred evaluation of the MIND system. All three
deliverables are due for PM 30, that is the end of the project.