The ELSNET Roadmap for Human Language Technologies

V1.0

DFKI GmbH

November 30, 2002

1 Contributors and Milestones

This is the final report of work supported by a grant of the European Community to the ELSNET-3 project. DFKI GmbH was subcontractor to Utrecht University to carry out the work reported here.

The principal investigator at DFKI GmbH was Prof. Dr. Hans Uszkoreit. Dr. Stephan Busemann has led the project. Thomas Eigner and Christian Federmann have carried out most of the implementation. Dr. Tania Avgustinova and Dr. Andreas Eisele have added work on the roadmap content.

The contractual milestones are described in this document in the following sections:

Milestones

1 *Summary documents of roadmap results in year 1, 2:* Section 5

2 *Proposal for a conceptual design of the roadmap:* Section 3.2.1

3 *Proposal for an interactive scheme to represent and maintain roadmap documents on the web:* for (re)presentation see Sections 3.2.2 and 3.2.3, for maintenance Section 6

4 *Implemented operational prototype of the scheme*: Section 4, see http://elsnet.dfki.de

5 *Inclusion of ELSNET3 year 1 and 2 roadmap results in website:* Section 5

6 *Public launch of the site and mechanisms to monitor discussion and elicitation of feedback:* Section 6

7 *Summary of year 3 roadmap results and integration in the website*: Section 5

8 *Final report*: This document.

The following issues need, according to this report, to be taken up or decided by the ELSNET Executive Board:

- Approval of Roadmap system and definition of a launch procedure. Page 23
- Who should participate in discussions on the Forum? Page 6, 11
- How can the results be promoted and disseminated? Page 19

2 Introduction

Human Language Technologies (HLT) are among the key information society technologies. They improve the human-computer interface, contribute to extracting useful information from an abundance of texts, support translation from one language to another, or summarize information according to a user's needs. In order to create contexts for joint efforts to face major challenges, roadmaps for HLT will help the community establishing common goals and priorities, and defining intermediate milestones.

For the first time, a public joint effort in creating a technology roadmap for language technologies has been initiated by the European Commission as part of the ELSNET network of excellence. As a first result of this ongoing work, a web site has been designed and implemented that presents the current version of the HLT Roadmap, which will be continuously updated (http://elsnet.dfki.de). Visitors can leave their comments and suggestions in the discussion Forum.

ELSNET is publishing the technology roadmap on the Web to serve as a major resource for the HLT community describing the state of the art and a vision of its future progress. This vision will continuously be extended, refined and modified by researchers and companies active in HLT. The ELSNET HLT Roadmap is also intended to form a scientific basis for decisions on funding research and development in the field.

Technology roadmaps are never finalized. Hence, the web page will always provide a state of information that demands correction, refinement, and update. The scientific and industrial community is invited to comment on the Roadmap and suggest extensions and modifications through public contributions to the moderated Roadmap Forum.

ELSNET will compile suggestions and create new versions of the Roadmap on this web site. Results from various roadmap workshops under the auspices of ELSNET will be included.

The ELSNET HLT Roadmap requires some technological prerequisites. It is easy to navigate. An on-line help menu offers short and useful explanations of the major features. Visitors will quickly discover how to access the Roadmap and how to provide comments, discussion contributions and feedback to improve it.

The Roadmap and the Forum are interlinked. The Roadmap Forum is moderated. Forum moderation as well as site maintenance are is organized through a set of password-protected site administration tools.

3 Conceptual Design of the HLT Roadmap Site

The concept of roadmap is a powerful and intuitive metaphor. A roadmap is a document that

- indicates directions for a planned journey
- shows how and in what order goals can be reached
- indicates distances
- is condensed in one structured presentation
- is perspicuously presented

A technology roadmap combines prediction of enabling developments, feasability judgements and scientific or economic goals into strategic planning. A technology roadmap puts these ingredients on a timeline, serving as a planning tool. This is feasible for periods of 5-6 years. Longer periods tend to include less reliable predictions.

3.1 Purpose and Scope

To design a technology roadmap, its purpose and scope need to be determined. The HLT community and industrial customers of that technology need to reach at a common understanding of the upcoming technological development in the field and of resulting applications. A ten-year period, starting with the first ELSNET Roadmap workshop in 2001, is envisaged. The common understanding is achieved by an interactive process of contributing and discussing on the Web. In principle everybody can participate. In practice, however, it might be advisable to restrict participation in the discussions. This needs to be determined by the ELSNET Executive Board. The scope of coverage needs to be identified and carefully delimited. Such delimitation allows for the definition of internal vs. external technologies that might enable some HLT result. External links may connect to existing or planned results or applications outside of the field of HLT.

3.2 Realization of Information

The realization of information in the Roadmap is defined on a structural, a representational and a presentational level. We use the following conceptual hierarchy.

3.2.1 Structural Level

Object Types

Areas and Subareas

Milestones

Applications

Relational Types

- Enabled by
- Enabling

- Timeline

- Scope
- Granularity

- Verbal Comments

- Comments on Milestones
- Comments on Times

- External Relations (people, concepts of other areas)

- Links to technology descriptions
- Links to commentators
- Links to enabling technologies

- 3.2.2 Representational Level

- Abstract Representation (e.g. XML definitions)
- Concrete Representation (Data Bases)
- Linking to Outside Sources

- 3.2.3 Presentational Level

- Conceptual Metaphor
- Visual Metaphor
- Types of Views
- Interactive Viewing
- Authoring, Editing, interactive Commenting
- Crosslinking with other knowledge sources

3.2.1 The structural level

The objects to be defined are milestones and relations between them. A milestone is a result – a technology, a tool, a resource, or an application. A milestone is characterized as follows:

milestone:= <unique identifier: int,
intuitive descriptor: string,
qualifications: setofquals,
achievement time: tempint,
source: docdesc,

comments: setofcomments>

Relations between milestones add to the temporal relationship inherent to the Roadmap. Obviously, for any milestone, enabling technologies and resources should be defined, as well as its role in enabling other technologies and applications. Other relationships must be definable if needed.

The timeline is defined on a yearly basis over a period of ten years.

Associating comments with the milestone enables discussion about any aspect of a milestone. It is possible to access all comments about a given milestone at once.

A mechanism supporting external links is provided. There are repositories – such as the ACL software registry, http://registry.dfki.de - and web portals – such as LT World, http://www.lt-world.org - available on the internet that can be used to access information about technologies, companies, key players, and resources.

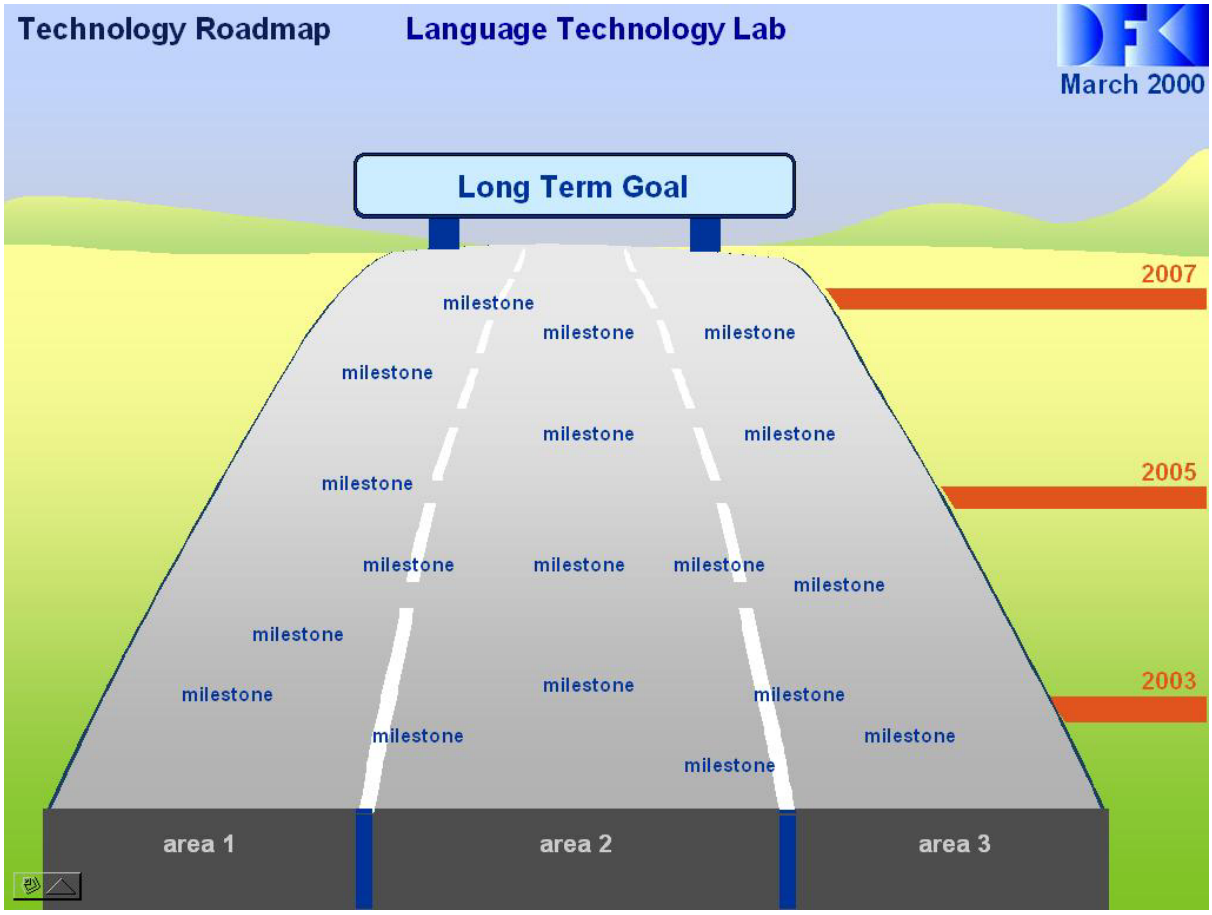

Figure 1: DFKI Technology Roadmap Scheme (Design: Hans Uszkoreit)

3.2.2 The representational level

The information types defined for the Roadmap are represented as typed objects in a database. The structure imposed on the database provides the relations between the objects.

Hyperlinks to outside sources are represented as marked textual objects in the database.

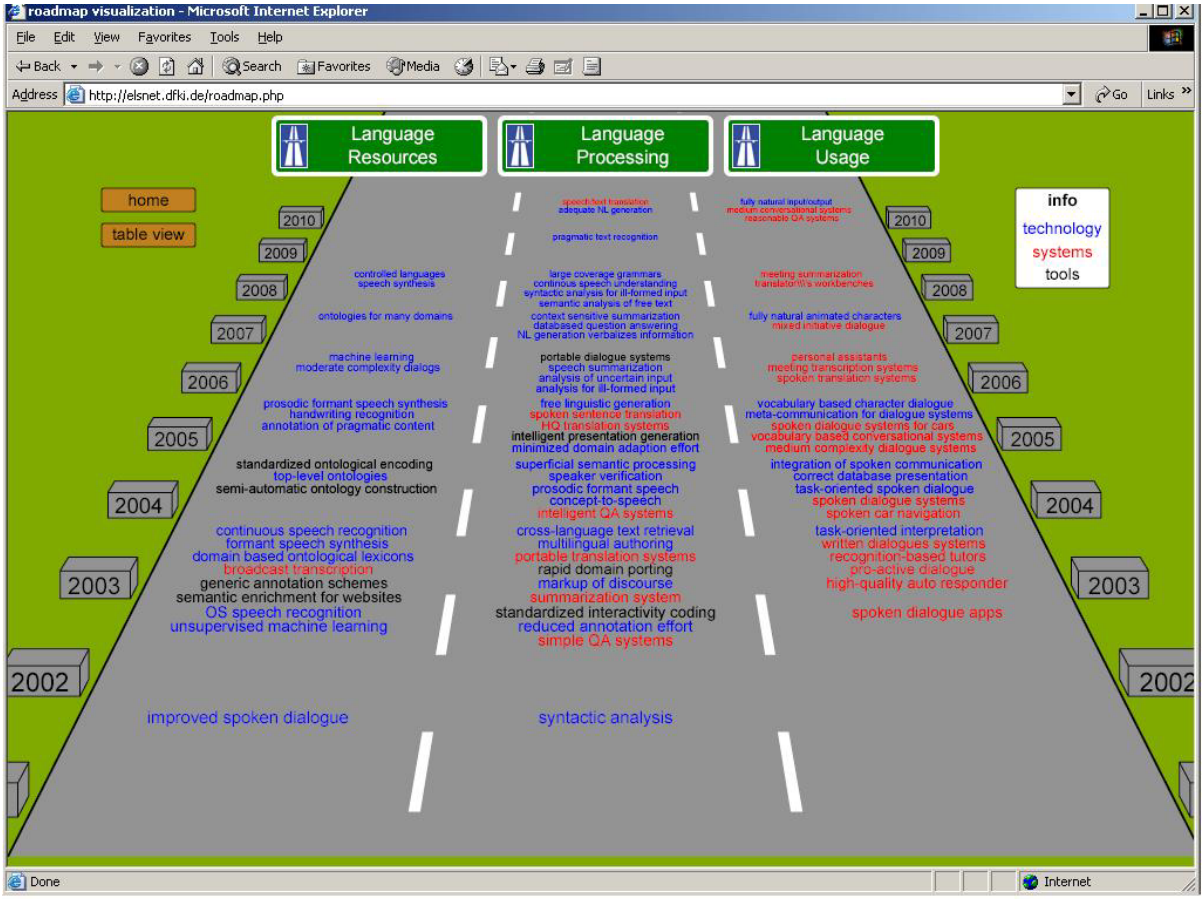

Figure 2: The Top-Level View of the ELSNET HLT Roadmap at http://elsnet.dfki.de

3.2.3 The presentational level

A web survey of various approaches to technology road mapping showed that a large variety of schemes is used to depict which type of technology will be available at what time. Many are two-dimensional functions of time and some functional values such as the throughput of electronic circuits. Others have graphical representations involving arrows, depicting successor relationship over time. Still others use a flow chart metaphor to depict the ways results feed into each other.

The roadmap metaphor has rarely been taken serious when it comes to graphical depiction. DFKI-internal roadmaps that have regularly been submitted to the board of shareholders for about ten years form a notable exception. The presentation consists of a three-dimensional graphical representation of a straight road leading slightly uphill, with distance marks depicting a three-, eight- or ten-years period (see Figure 1). The road is shown from a driver's point of view, from a slightly elevated angle. The major milestones are placed on the road surface. They fall into different categories that are marked by colours. The lanes correspond to the main areas of technology. Lane directions are shown on a highway sign spanning the entire road.

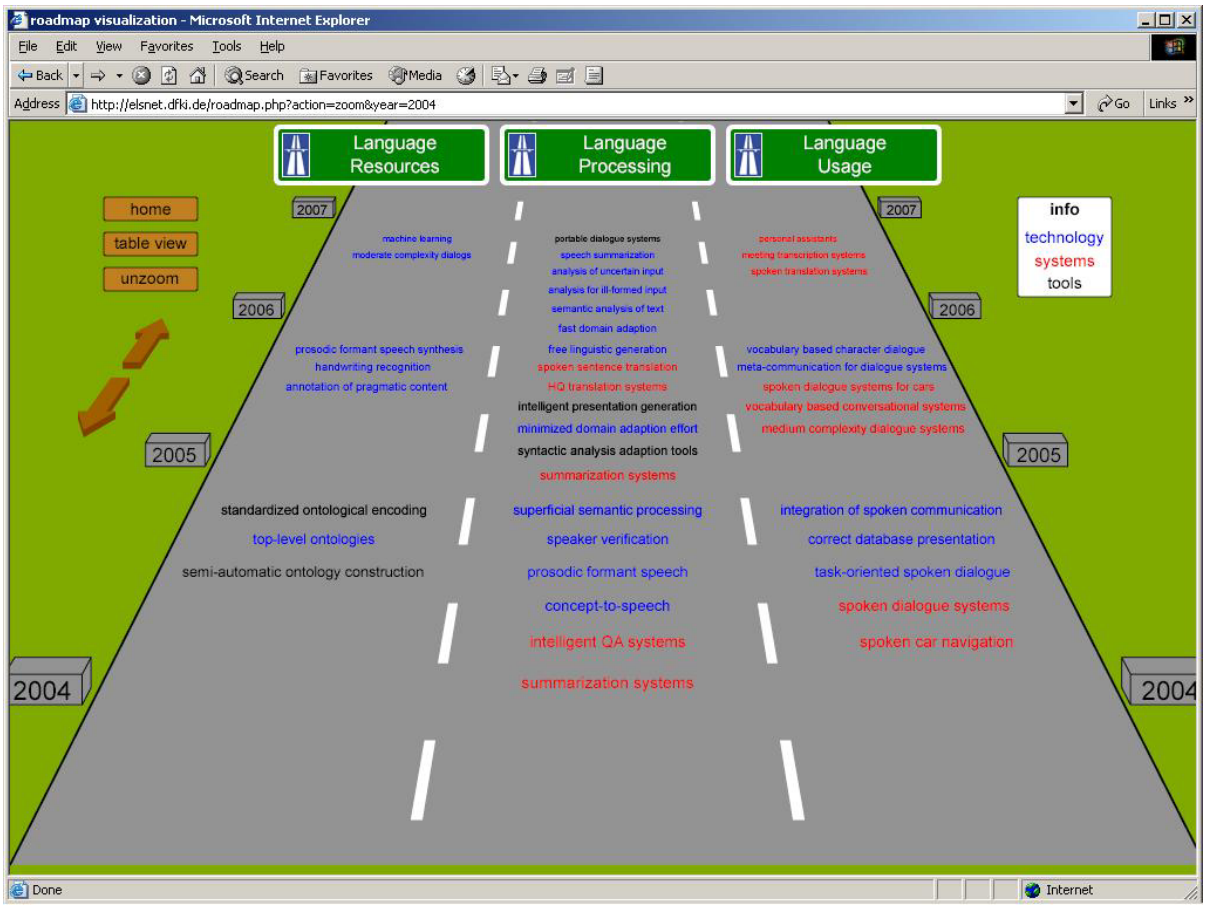

Figure 3: A Three Years Period at the Zoom Level of the ELSNET HLT Roadmap

This presentation has been refined and taken over as the top-level presentation of the ELSNET HLT Roadmap (Figure 2). Besides, a tabular presentation is accessible (button “table view”) that maintains the time axis by listing the milestones in their temporal order in three rows corresponding to the three lanes.

One of the natural visual features of the roadmap metaphor is the relative difficulty with which more distant milestones, as opposed to near ones, are recognized. This corresponds to the fact that technologies in the far future are less precisely defined a less predictable than those in the near future. But forecasting requires us to mentally travel into the future; zooming along the time axis mimics this. The second-level visualization focuses on any three years period, as shown in Figure 3. The brown arrows allow the user to shift that three-year selection by one year in either direction. Selected areas of any three-years period can be enlarged (see Figure 4).

At any presentational level shown so far, a menu with options for details, comments, and relations can be opened for any milestone.

The “view details” option (see Figure 5) provides the short name, as seen on the road surface, the full description for the milestone, which can not be presented on the road's surface due to space limitations, and the category. The details view also allows for linking to HLT resources such as LT World (http://www.lt-world.org). Another slot can be used to show additional information, text or link, related to the milestone.

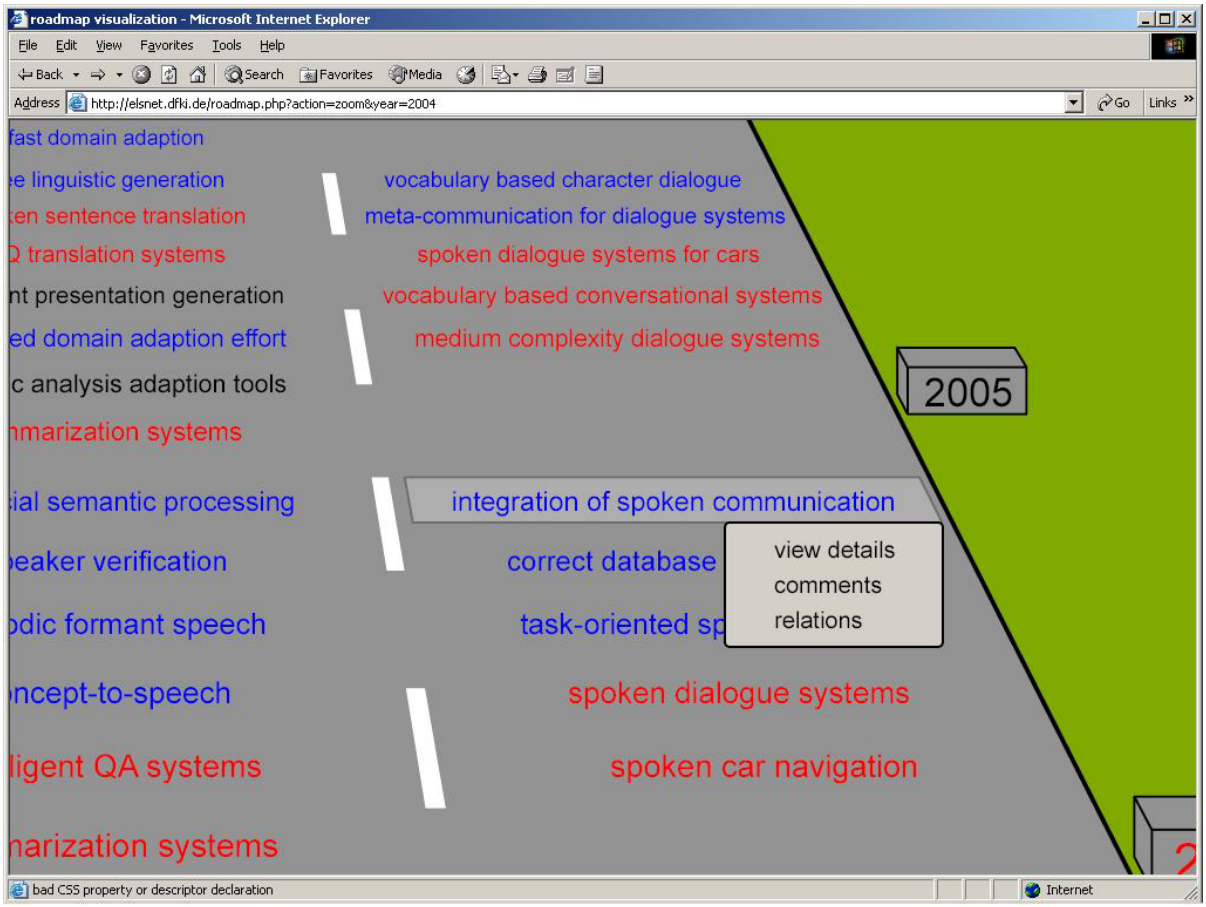

Figure 4: Enlarged Selection with Menu Shown for One Milestone

The “comments” option takes care of interactivity. It allows the user to send a comment about this or any other milestone to the Roadmap Forum. The user can reply to existing messages in the Forum or open up a new thread. The message will show up in the Forum after a moderator approves it[1]. A decision about who will be admitted to the Forum needs to be taken by the ELSNET Executive Board. The Forum can be accessed from the Roadmap, but it is equally possible to access the Roadmap from the Forum.

[1] The Forum will be moderated by DFKI GmbH as part of the ELSNET4 contract.

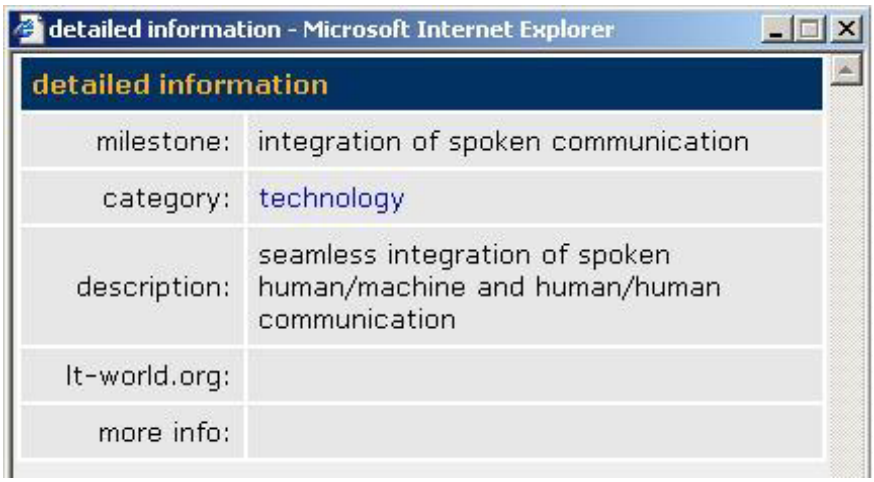

Figure 5: Opening the “detailed view” Menu Entry

The “relations” option presents dependencies between milestones. This feature has gained in importance in recent discussions of the ELSNET Executive Board about the Roadmap system. Only by establishing relationships between milestones it is possible to express more than simple temporal subsequence.

Relations can be defined freely and are visualized in different colours. For a milestone, the “relations” option displays all relations associated with that milestone. All other milestones, which are not involved in the relations, are shadowed. The name of the relation is visualized when the mouse is hovering above the link.

In

Figure 6, the milestone “meta communication for dialogue systems” has two relation instances associated with it. They represent the consideration that the technology described by the milestone is enabled by “speaker identification” and “spoken dialogue systems”.

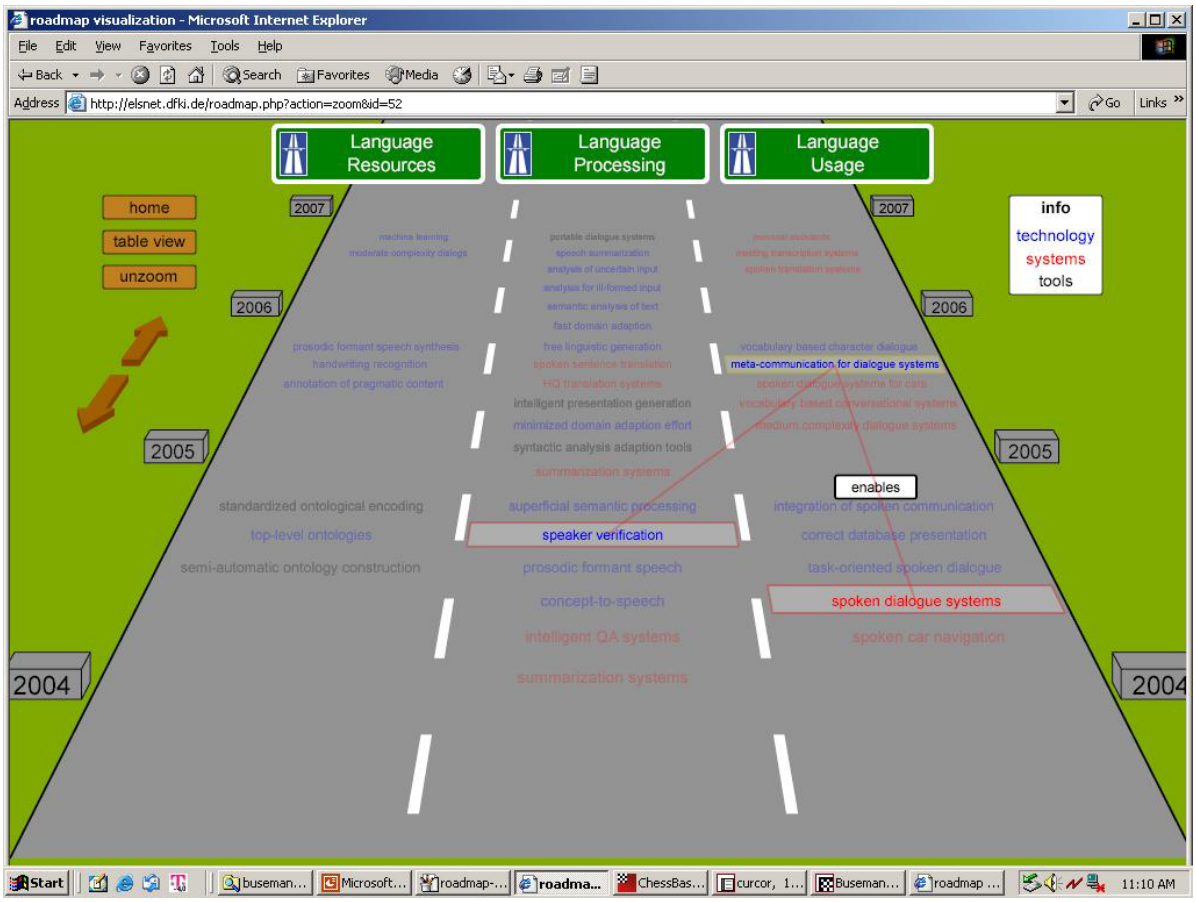

Figure 6: Displaying Relations Between Milestones

4 Implementation

The ELSNET HLT Roadmap has been implemented using scalable vector graphics (SVG) as provided by Adobe. The Roadmap objects are implemented using a MySQL database. The website http://elsnet.dfki.de, running on an Apache web server, is created automatically from the database using PHP scripts.

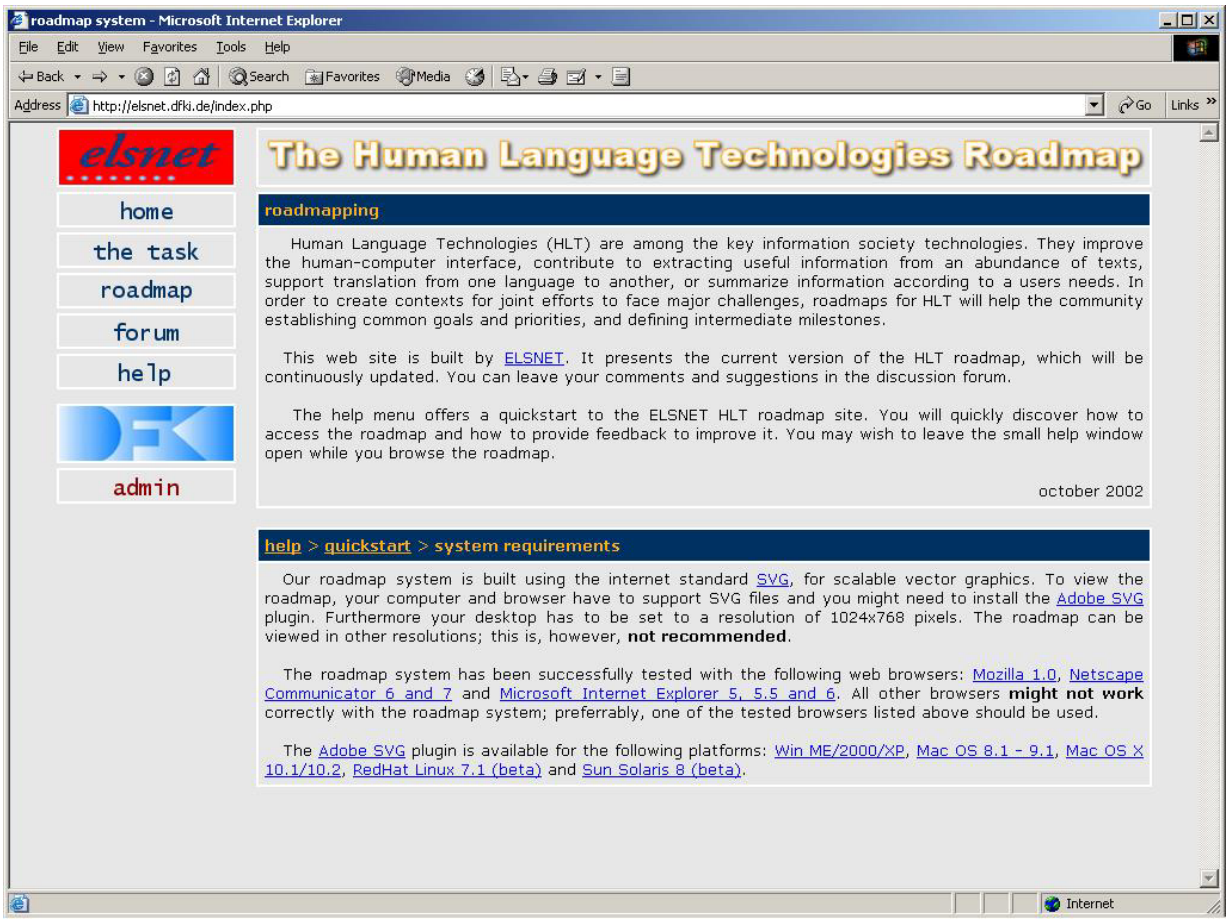

Figure 7: The HLT Roadmap Web Site at http://elsnet.dfki.de

4.1 Technical Prerequisites

The Roadmap system is built using the Internet standard SVG, for scalable vector graphics. To view the Roadmap, the computer and browser have to support SVG files and the Adobe SVG plug in must be installed. Furthermore the desktop has to be set to a resolution of 1024x768 pixels. The Roadmap can be viewed in other resolutions; this is, however, not recommended.

The Roadmap system has been successfully tested with the following web browsers: Mozilla 1.0, Netscape Communicator 6 and 7 and Microsoft Internet Explorer 5, 5.5 and 6. All other browsers might not work correctly with the Roadmap system; preferably, one of the tested browsers listed above should be used.

The Adobe SVG plug in is available for the following platforms: Win ME/2000/XP, Mac OS 8.1 - 9.1, Mac OS X 10.1/10.2, RedHat Linux 7.1 (beta) and Sun Solaris 8 (beta).

4.2 The Structure of the Web Site

Figure 7 shows the entry page of the ELSNET HLT Roadmap system. On the left-hand side, a set of buttons is visible (in fact, the buttons are available on all Roadmap system pages except those displaying the 3D Roadmap itself). The buttons provide the following functionality:

- 'home' leads to the page shown in Figure 7.
- 'the task' describes the mission statement, which is partly repeated in Section 2; it also describes the current version (see Section 5).
- 'roadmap' leads to the top-level view of the HLT Roadmap (cf. Figure 2).
- 'forum' leads into the discussion forum, where new threads can be started, or existing be extended on. At the time of the launch, the Forum is empty.
- There is an online help facility available ('help') that includes a 'quick start' while surfing the Roadmap site, and a description of the 'Roadmap and forum' aiming at a more thorough explanation.
- 'admin' is a password-protected area for the system administrators. The admin tools offer access to the database. Any modifications to the contents and also many ways of modifying the structure of the web site can be realized here. See Section 6 for details.

5 Contents

The content that is currently implemented is discussed. Methods for producing new content and for consensus building are presented

5.1 Current Content and Discussion

The current version 1.0 is based on the documents that compile results from workshops in the first and second year of the project:

- Nils Ole Bernsen: *Speech-Related Technologies: Where will the field go in 10 years?* ELSNET Roadmap workshop at Katwijk, The Netherlands, November 2000.
- Andreas Eisele and Dorothea Ziegler-Eisele (eds.): *Towards a Roadmap on Human Language Technology: Natural Language Processing.* March 2002. Report of results achieved at the ELSNET Roadmap Workshop in Toulouse, France, July 2001.[2]

Implementing the milestones suggested in these papers using the suggested Roadmap scheme required some additions to the content of the papers. In particular, the authors were not as keen as to assign a date to every milestone. Dates have been elicitated and added. The short names on the road have been derived from the longer descriptions taken from the papers. Moreover, the lanes and their semantics (see Figure 2) as well as the classification of the milestones into technologies, systems, and tools were added. Finally, a few relations of type 'enables' have been included for demonstrational purposes.

Links into the LT World have so far not been provided for two reasons. The nomenclature used in the LT World and that by the contents currently implemented in the Roadmap system differ much, so that an attempt at an automatic linking was unsuccessful. Defining and automating a link mapping is thus left to future work (cf. Section 7). The contents of version 1.0 has so far not been discussed and approved as suitable for a WWW-based technology roadmap. Justified doubts were raised about whether a set of milestones would be the best possible starting point for a public discussion.[3] Hence a labour-intensive manual link-up would have run the risk of not paying off in the course of on-going revisions.

As was noted in this discussion, the work documented so far is indicative in the sense that no full definition of the notion neither of milestone nor of what it means for a milestone to be available in a particular year is provided yet. The section on qualifications of Bernsen's paper addresses this issue.

The Board also suggested adopting applications as milestones. It would be preferable from a communication point of view to aim at forecasting challenging applications and identify their enabling technologies. This would make it easier to attract valuable discussion of the Roadmap content, since pure technology forecasting would be left to a small group of experts.

---

[2] This document is part of the present subcontract (Milestone 1, cf. Section 1).

[3] Meeting of the ELSNET Executive Board in Berlin on October 8th, 2002.

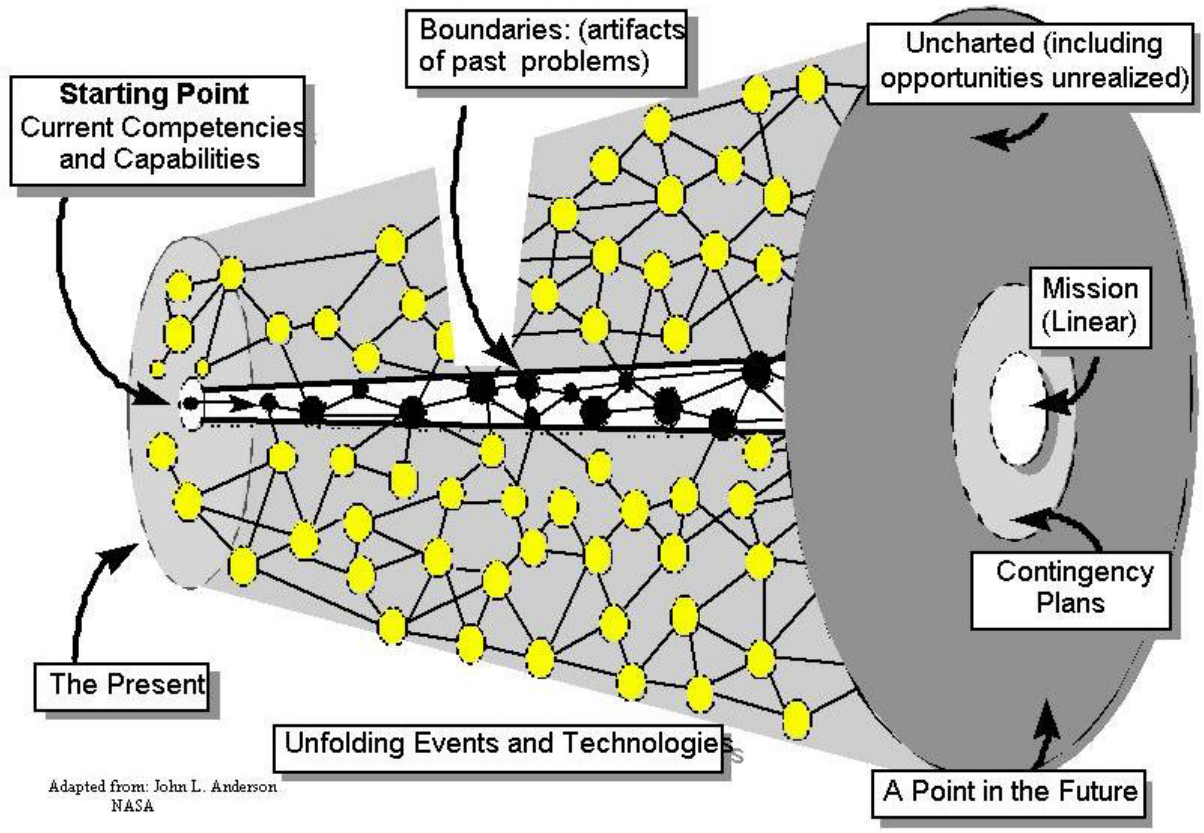

Figure 8: Linear Extrapolation: The Corporate Mission

5.2 Producing New Contents

This section details suggestions relevant for the usage of the HLT Roadmap.

Based on the input, the Roadmap will be corrected, augmented and adjusted by the administrator.

5.2.1 Consensus documentation

We identify methods of documenting a consensus reached (cf. also Section 5.4).

Intermediate results are published in the Roadmap. Potentially, more than one roadmap must be provided to document alternative views. The discussion on milestones in the Forum is basically open to everyone. Comments and endorsements are solicited. The publication of subsequent editions document the progress achieved. Summaries of comments on previous versions are integrated. Feedback is gathered at professional meetings.

5.2.2 Resources

A variety of results can be gathered, exploited and fed into some consensus building process:

- Gather specialized input (area roadmaps) from past and future workshops in machine translation and multimodal interactive systems;

- Monitor relevant other areas, especially enabling technologies such as mobile telephony (UMTS, interfaces), INTERNET 2, VoiceXML, electronic books, electronic paper.

5.3 Speculative Views

Most of the times, the future turns out different from earlier predictions. A major reason consists in failure to consider alternatives. Paths seeming slightly less likely from today‘s perspective are not explored. Dependencies from external factors are not taken serious.

As a remedy, some relevant dependencies on enabling developments and funding support need to be identified and discussed. Small „What if...“ discussion sessions may provide new insights and assessments.

Instruments are known that can support this approach. The Horizon Mission Methodology (HMM) provides a systematic way for users to step outside the bounds of the 'flashlight beam' (cf. Figure 8) to "see", think, and plan. The HMM was developed since 1993 at NASA by John Anderson et al. The method consists of five steps:

1. Make an intuitive leap outside the bounds of the 'flashlight beam" to create extraordinary, impossible, ‘mind-blowing' alternative futures - the New Horizons - those beyond current reach. The New Horizon should be strategically relevant and plausible, but it must also leapfrog foreseeable solutions.
2. Construct a new frame of reference by defining the New Horizon in terms of: motivating forces, novel, unique and/or critical functions, pivotal or unprecedented capabilities and extreme performance levels. In other words, invent the 'world' or create the vision to be brought about by this New Horizon. This world should include assumptions about economic, technological, cultural and political attributes and drivers.
3. Think within this new frame of reference to identify breakthrough innovations (in both activities and technologies) required to achieve that future. The key is that functional capability and engineering alternatives should be expressed at a higher level than simple incremental activity or technology alternatives. This step is more difficult than it seems. It requires considerable discipline to stay on track - to think first at a higher functional level than simply drive to a ‘‘right technology answer'.
4. Begin the return 'back from the future' to the present. The higher-level functions can then be clustered into vision-inspiring categories. Evocative metaphors are extremely productive and integrative at this stage of the process. High leverage categories such as entirely new capabilities, new common (or multi-use) technologies, infrastructure changes, new uses, applications, and dramatic potential payoffs can then be identified and explored.
5. The high leverage concepts are then transformed back into the present by identifying and working through the business value and technology value chains and relating functional activities and selected technologies to functional activities (cf. Figure 9). Near-term steps and issues such as: new products,

new markets, new capabilities, investment and required breakthroughs (e.g., Technologies), other drivers and motivating forces and insights concerning competitive activity in these areas can then be identified and programmed for pursuit and response after being identified during environmental scanning.

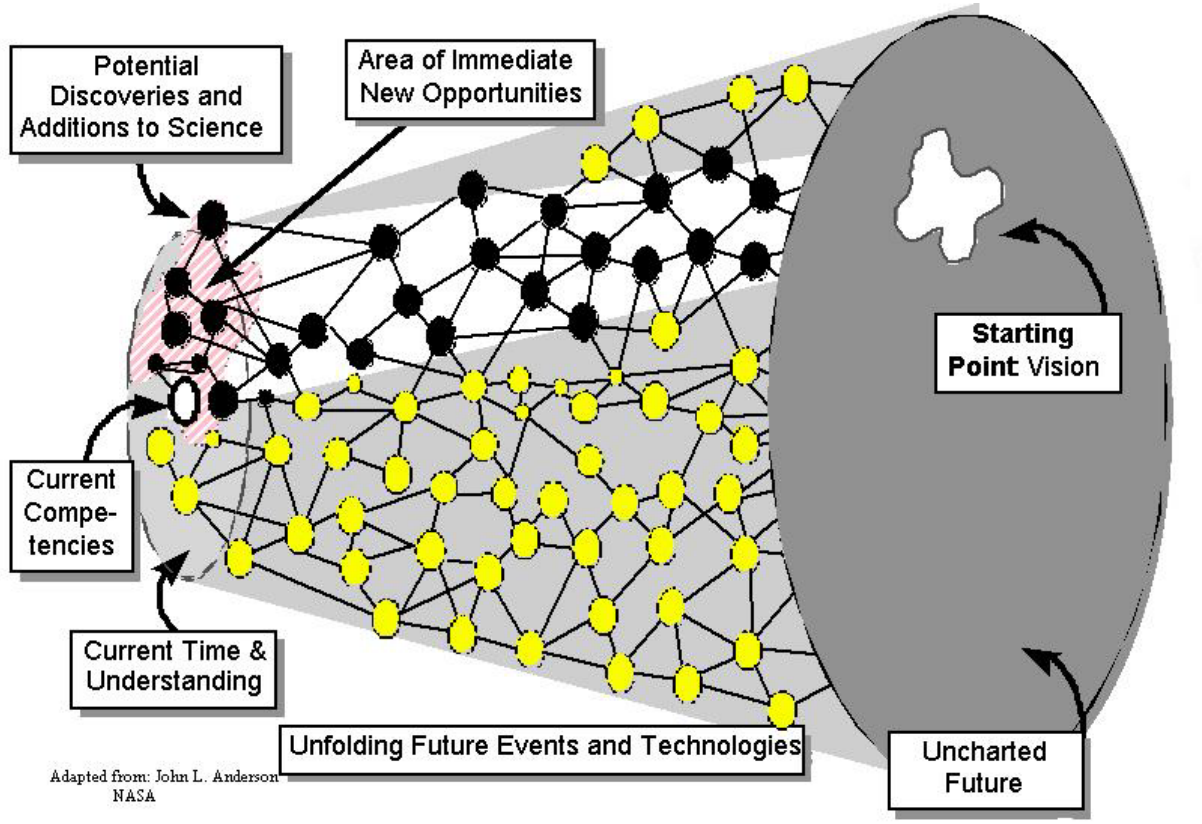

Figure 9: “New Horizons”: Managing the Present From the Future.

5.4 Exploitation

It is obvious, that any changes to the Roadmap content should not rely within the responsibility of the administrator. The decision-making process needs to be defined that eventually leads to a revision of the Roadmap.

Promotion, dissemination to other technological and social sectors, as well as channelling into advisory and decision processes are tasks to be performed by ELSNET, to be planned by the Executive Board.

6 Maintenance

The Roadmap system requires two special user roles for maintenance. The *moderator* decides which messages sent to the system will be published in the Forum. She should act within 48 hours, at times much faster, to ensure an on-going flow of the discussion. The *administrator* changes the contents and part of the structure of the Roadmap, using the admin tools. These roles are separate from that of the *programmer* who would be involved when database functionality was at stake, or the general look and feel of the Roadmap system was to be changed.

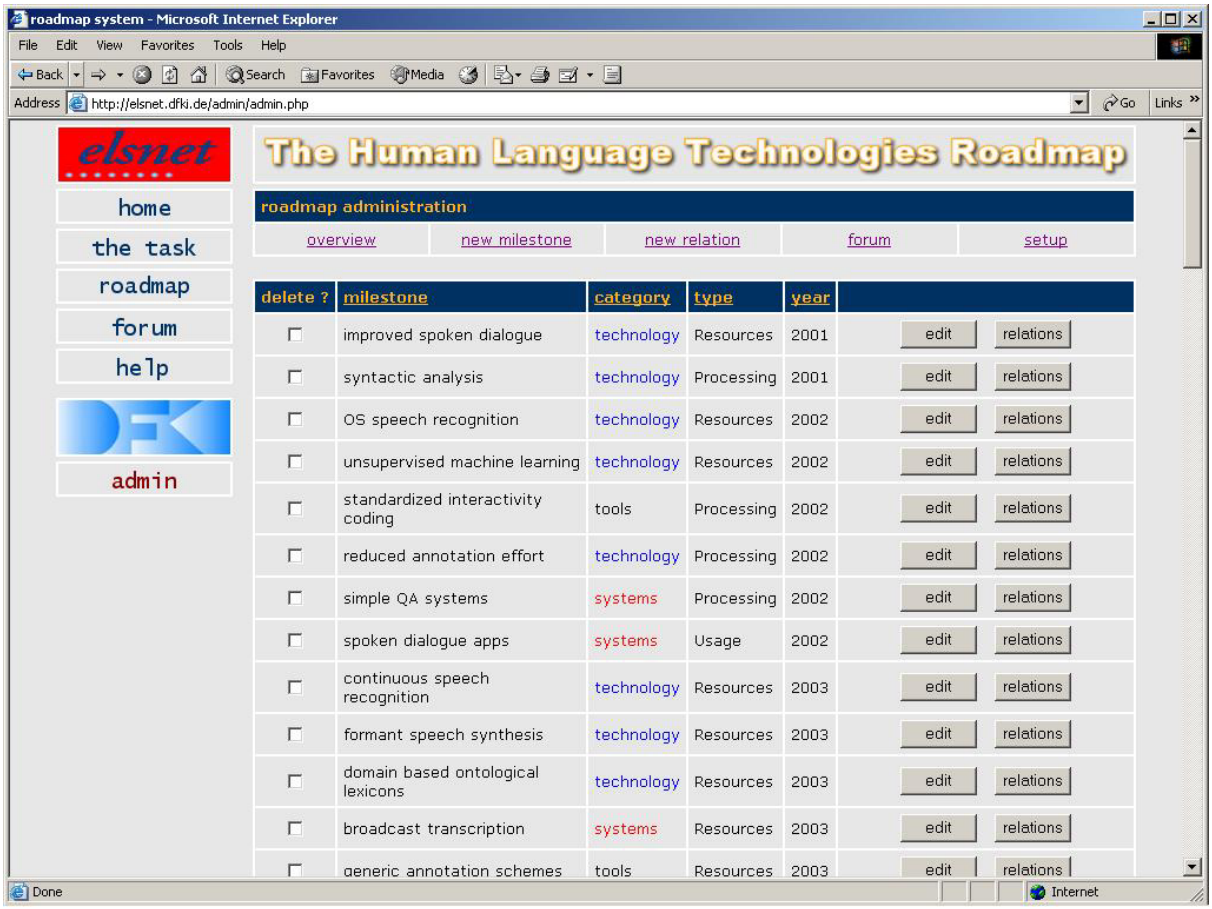

Figure 10: Top-Level Page of the Admin-Tools

6.1 The Administrator's Role

The administrator has various tools at her disposal, which will be presented in the sequel.

When entering the password-protected 'admin' section of the HLT Roadmap system, the top-level page in Figure 10 is displayed. On the top, several links are offered that persist on all admin pages:

- 'overview' points to the top-level page (Figure 10);

- ‘new milestone’ links to the tool for defining and editing milestones (Figure 11);
- ‘new relation’ links to the tool for defining and editing relations (Figure 12);
- ‘forum’ links to the tool for monitoring the discussion Forum, which is the main task for the moderator;
- ‘setup’ links to the tool for defining and changing features of visualizing milestones and relations (Figure 13).

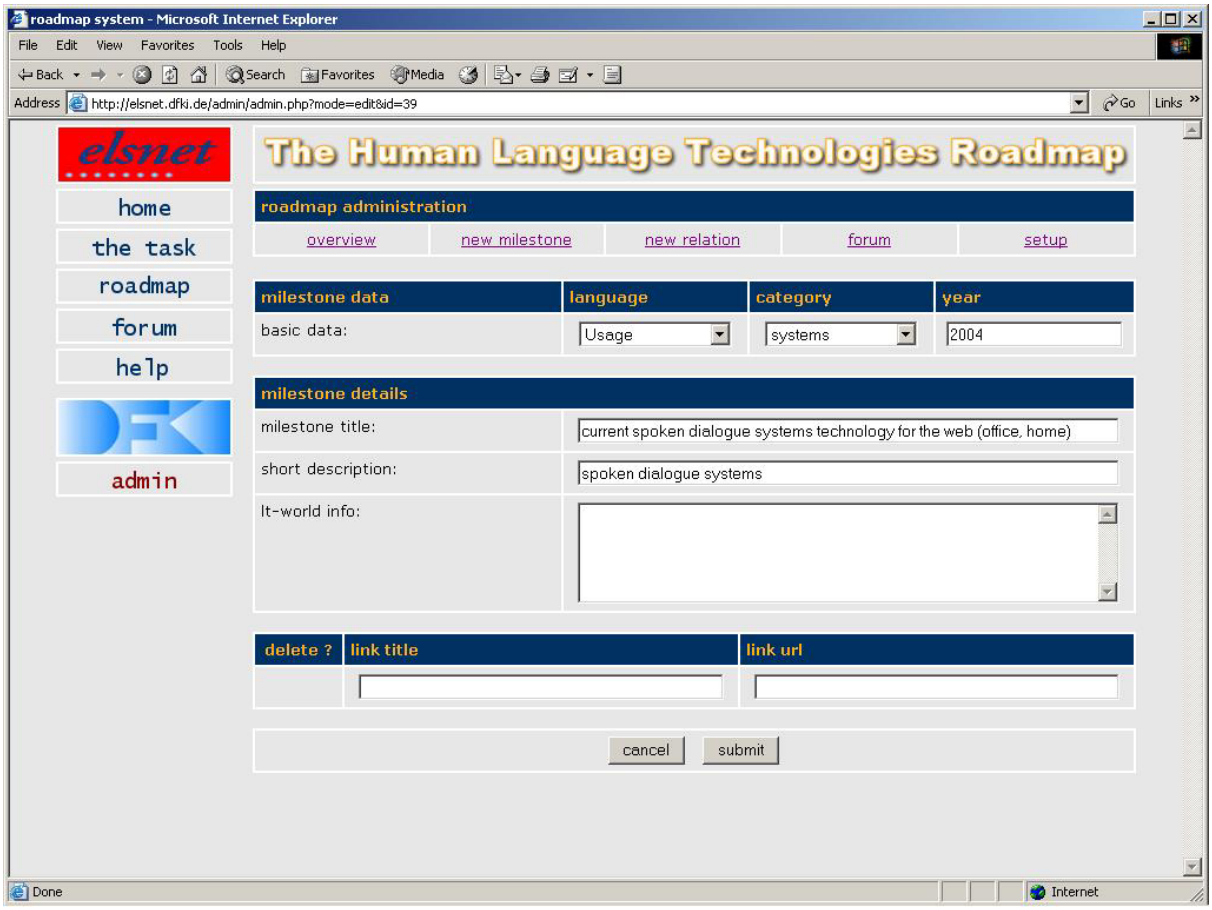

Figure 11: Defining or Editing Milestones

The top-level page displays all milestones defined in the database. They can be sorted according to their alphabetical order, the category, the type, or the year. The tool allows for deleting or editing a milestone as well as for placing it into a relationship with some other milestone.

Milestones are defined or edited using the interface in Figure 11. The basic data consist of type, category and year. The milestone details include the title, which is the description displayed on the road, and a short description that defines the milestone more precisely. Links to external resources, such as LT World, are defined here as well. The milestone details defined here can be retrieved from the Roadmap by investigating the detailed information of the milestone.

Relations are defined or edited using the interface in Figure 12. The type of relations and the participating milestones are selected. For convenience, the milestones can

directly be accessed from the tabular view underneath the menus in order to get the necessary information.

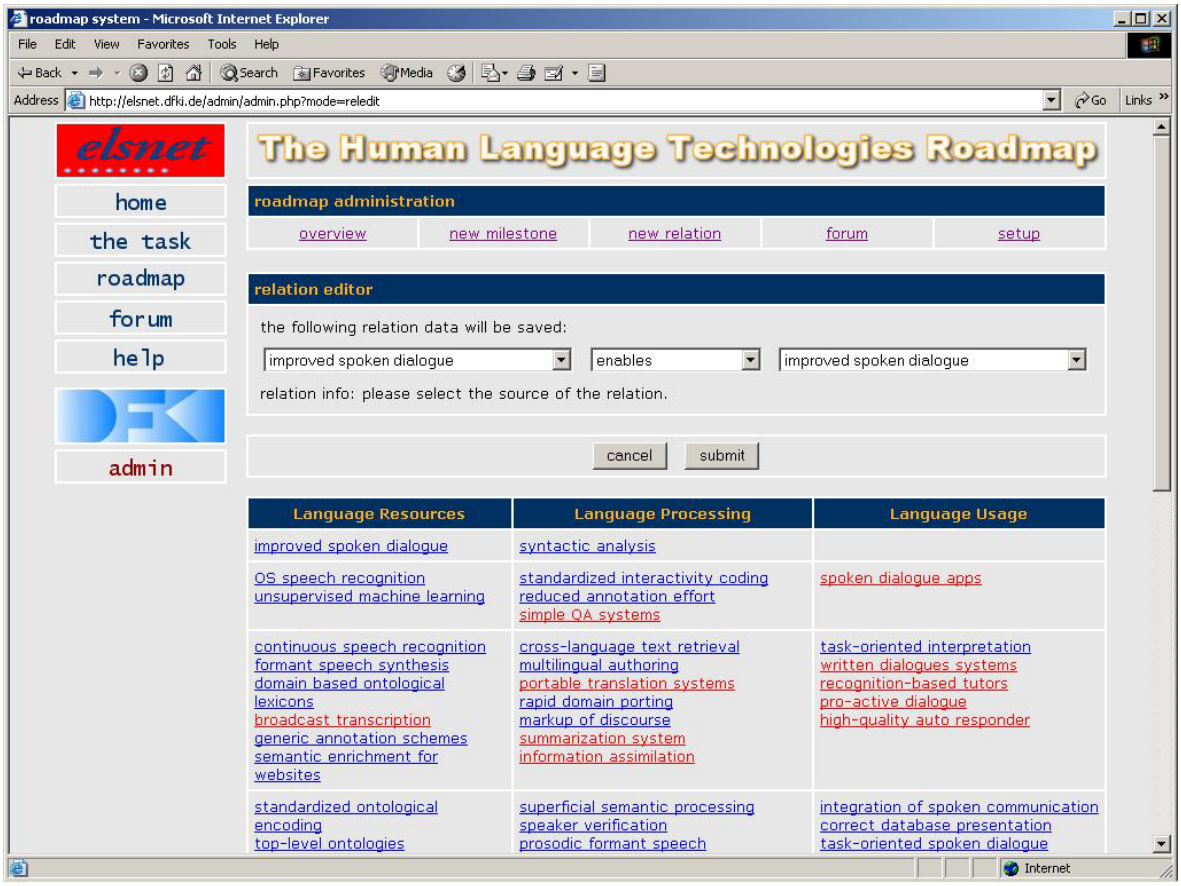

Figure 12: Defining or Editing Relations Between Milestones

The type of relations is defined using the tool shown in Figure 13. In the lower half, up to three different relation types can be distinguished by a name, a short description and a colour of the link between the arguments in the Roadmap.

Other presentational features of milestones can also be influence on this page. The categories, the types and the colours of milestones can be modified.

6.2 The Moderator's Role

The moderator will concentrate on the Forum. She will examine the list of incoming messages and submit the accepted ones for display in the Forum. It will be entirely in her responsibility to accept or deny messages. The moderator will report to the ELSNET Executive Board. It is advisable to design and publish guidelines, thus making the usage rules transparent.

6.3 Public Launch of the ELSNET HLT Roadmap Site

The Roadmap site will be launched after the ELSNET Executive Board has approved it. The procedure of the launch needs to be clarified, including decisions about dedicated publications in major HLT-related media. The launch has to be initiated

from ELSNET. DFKI will support it from a technical point of view and with the activities committed to as a project partner of ELSNET4.

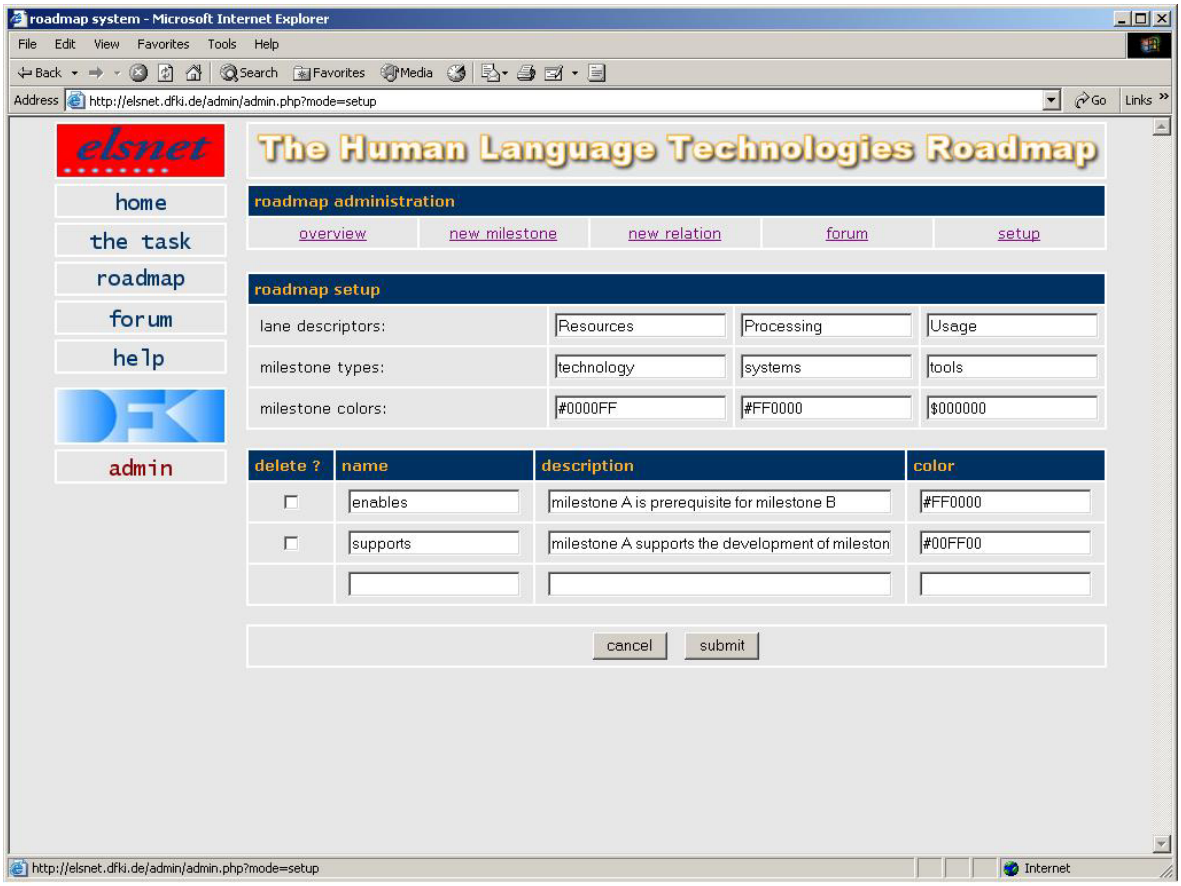

Figure 13: Defining or Modifying Presentation Features

7 Future Work

Several open ends require further work, to be carried out in the framework of the ELSNET4 project.

7.1 Linking to External Sources

As mentioned before, automatic hyperlinking of milestones has not been investigated yet thoroughly. The LT World offers an extensive knowledge about existing language technologies that form part of innovative applications or enable new technologies. Links from corresponding milestones to definitions, authors and researchers connected to existing technologies should be generated automatically, as the Roadmap is subject to changes and manual linking is labour-intensive.

The main difficulty for an automatic approach consists in the different nomenclature between the contents implemented and the definitions in the LT World. While this is telling in itself - even experts in the field don't seem to talk in languages that match up top each other – it requires the definition of an ontology for a successful mapping. Such an ontology is currently being developed as part of the DFKI project COLLATE.

7.2 Visualizing Relationships

Browsing the Roadmap is currently only possible at the top level, which corresponds to a ten years period (cf. Figure 2), and on the zoom level for a three years period (cf. Figure 3). Any other zooming using SVG does only enlarge a selection at the respective level, but does not access additional portions of the Roadmap.

Relations can only be accessed from within the zoom level. This makes it impossible to visualize relation definitions that extend over more than three years. For the visualization of arbitrary relationships, any selection of the Roadmap can be relevant. The next version of the system will therefore allow for more flexibility with selecting portions of the Roadmap.

7.3 Improving Forum Presentation

When new messages are added to the discussion Forum, they are characterized by the selected milestone or theme. There is currently no distinction on whether a message was sent as a reply to another one, or just it is yet another comment on the same theme.

It has been considered to mark reply messages by using a tree structure that could be represented by a suitable indentation of message headings. Experience with the Forum will show the need to explicitly trace discussion threads. Other discussion Forums offering this functionality demonstrate, however, that it is by and large up to the participants to use this option in such a way that it is helpful to others (see e.g. http://www.correspondencechess.com/bbs/, "The Correspondence Chess Message Board").