

TextGrid as a Digital Ecosystem

Marc Wilhelm Küster*, Christoph Ludwig* and Andreas Aschenbrenner†

* University of Applied Sciences Worms, Erenburgerstr. 19, 67549 Worms, Germany, Email: {kuester|ludwig}@fh-worms.de

† Goettingen State and University Library, Papendiek 14, 37073 Goettingen, Germany,
Email: {aschenbrenner}@sub.uni-goettingen.de

Abstract—The TextGrid project brings together eight German institutions from both academia and the commercial sector to “create a community grid for the collaborative editing, annotation, analysis and publication of specialist texts”. Leveraging the grid infrastructure, textual data and supporting images from various research projects and content providers such as archives and libraries can fuse into a virtual corpus that can be seamlessly searched and analyzed.

TextGrid will be a key enabler for collaborative textual scholarship, aiming at overcoming current isolation in research and facilitating cooperative working methods and the sharing of resources, content and software agents alike. It will also enable quantitative and comparative studies across corpora on a scale that might otherwise have been impossible to achieve.

I. THE LONE SCHOLAR IN HIS ATTIC

For more than four centuries now the ideal of textual scholarship is all too often that of individual intellects working with great erudition, perseverance and dedication on their chosen subjects in isolation. Their projects could easily last for many years if not decades without any substantial interaction with their peers. Quite frequently death struck before they could complete their work, obsoleting most or all of their intermediate results.

A stunning, but by no means atypical example for this is the history of the edition of the so-called pseudo-capitulars supposedly written by a Benedictus Levita. This grandiose forgery from the middle of the ninth century AD culled its wisdom from a large number of sources and presented them as legal precedent (in the widest sense) in 1732 unordered chapters in three books and four *additiones*.

Even though a historic source of first order, this massive text has last been successfully edited by the French scholar Étienne Baluze in 1677 after a first attempt by Jean du Tillet in 1548. Unsurprisingly, this aged edition, by now well over three hundred years old, is long considered to be outdated. Yet the text has defied full scholarly edition since, as [1] relates, on which this case story is based. In the last century alone a sequence of no less than four scholars attempted a critical edition, three of them dying before completion, forcing their successors to (almost) start again from square one. Only the fourth team now — using modern technology and collaborative working methods — stands a reasonable chance of actually completing this mammoth task.

That such a procedure is wasteful in the extreme and not adapted to today’s rhythm of research needs no further

exposition. We need to replace a culture of isolated individuals by one of collaboration and sharing [2]. This problem is as much an organizational as a technical one. On the technical level, we need a platform that allows communication and data sharing across organizational, geographical and disciplinary boundaries while offering the necessary tools to automate much of the drudgery of textual criticism and analysis. On an organizational level, we must bring together players of quite diverse origins and interests — domain experts such as philologists, historians, and linguists, content providers such as archives and libraries, and technical experts providing the infrastructure — who have not much of a record of collaborating with each other.

II. THE TEXTGRID VISION

TextGrid wants to become an answer to these problems.

Partially funded by the German Federal Ministry of Education and Research (BMBF) under the D-Grid initiative by agreement 07TG01A-H¹ for a three years period starting in February 2006, TextGrid brings together eight German institutions from both academia and the commercial sector to “create a community grid for the collaborative editing, annotation, analysis and publication of specialist texts” [3]. Leveraging a grid infrastructure, textual data and supporting images from various research projects and content providers such as archives and libraries can fuse into a virtual corpus that can be seamlessly searched and analyzed.

Participating organizations provide dedicated services — software agents — with well-defined interfaces that can be harnessed together through a user defined workflow to mine or analyze existing textual data or to structure new data both manually and automatically. New agents can be developed, deployed and integrated easily.

From this soil organizations and agents together will eventually form the nucleus for a new eHumanities ecosystem that transcends the current solipsism.

III. THE TEXTGRID PROJECT

Over the last years, many organizations and projects started digitizing archives and corpora. The collective data volume of all these texts, markups, and annotations as well as the digital facsimiles exceeds multiple terabytes in Germany alone.

¹Responsibility for the contents of this publication rests with its authors.

Software systems like TextGrid that intend to integrate (from a user's perspective, at least) such huge, distributed data repositories need to adopt special strategies for data storage, retrieval, and maybe relocation. Collaboration and interaction among the users make access control, authentication, and authorization more challenging and introduce additional synchronization issues. Processing of large corpora can consume significant compute resources and requires a sophisticated scheduling mechanism. Grid computing has set out to provide solutions for these problems, whence TextGrid was started as one of initially six community grid projects (CG) within the D-Grid program [4].

D-Grid was set up as a long-term strategic program to establish a national Grid infrastructure in Germany. Being part of the BMBF's eScience initiative, its goal is to provide basic, sustainable resources and services that other eScience projects can easily build upon. Services and modules developed in those projects shall be made available to other projects if they are generic enough. Since more than 100 institutions belong to D-Grid, this is already a highly collaborative program.

The six D-Grid CGs plan to leverage the Grid technology for quite heterogeneous disciplines: From high energy physics over engineering and medical research to the humanities. Researchers from such diverse areas meet in D-Grid workshops where often a considerable overlap in the infrastructure requirements of their respective CGs is identified. Thus, D-Grid fosters cross-disciplinary communication and collaboration; partners from TextGrid and other CGs already proposed and started joint D-Grid infrastructure projects.

Cross-disciplinary collaboration is also practiced in TextGrid itself. Among the TextGrid partners — State and University Library (SUB) Göttingen as project coordinator, Darmstadt University of Technology, Institut für Deutsche Sprache in Mannheim, University of Trier, University of Applied Sciences Worms, University of Würzburg, and our commercial partners DAASI International and Saphor, both in Tübingen — are domain experts for critical edition, lexicography, linguistics, as well as computer scientists.

Besides the technical expertise, the TextGrid partners bring the different research goals that come with their respective special field into the project. They work on different types of texts or text corpora, they encode different aspects of their texts, and they analyze their data with quite different questions in mind. For example, linguists may tag the grammatical form of single words in the text, whereas philologists may encode in which revision the author (or the publisher, a scribe, or whoever) added or deleted text parts of arbitrary length.

Nevertheless, the tools the researchers apply are often similar or the same: For example, all historic manuscripts need to be digitized, transcribed, and encoded. If there are several editions of a text or we have the same text in several manuscripts, then a collation tool has to assist with finding the differences. After these steps are applied and possibly refined in iterations with human control and feedback, other, more domain specific tools may be used.

TextGrid aims to collect all these tools as services on a

common platform. The more generic services will be put to use by most of the users, while others are of interest to a smaller audience only. But all services can be combined in arbitrary workflows; the availability of services from neighbor domains of textual research may even lead to new applications of existing tools. TextGrid will therefore build a platform that connects experts and enables research in related fields. The project partners identified in their project proposal a subset of text processing services that need to be available early on:

- An XML editor that validates documents according to DTD, RelaxNG or XML Schema and that can be customized by suitable style sheets for the task at hand.
- A metadata editor that supports freely configurable metadata structures and validates the entered data.
- Editors for creating links from and to both text and manuscript fragments.
- A tool to handle bibliographic references.
- A tokenizer customizable for different languages.
- A lemmatizer that relies on selectable lemma lists and provides for manual correction. Lemma lists for several historic forms of German have to be available as well.
- A collation service that compares an arbitrary number of variants of a text document and encodes the differences found according to guidelines specified by TEI [5].
- A sorting service that respects locale-dependent or user defined rules.
- A streaming editor that transforms text documents according to a given rule set. Neither input nor output are necessarily XML documents, even though we expect this to be the most common case.
- A service to interface OCR software.

Besides these services more or less specific to textual research, further services for finding and retrieving data in the grid, for publishing, for the generation and execution of workflows etc. are also required, of course. TextGrid will evaluate existing tools used for these tasks, adapt and integrate them where reasonable or re-develop them where necessary.

The TextGrid services need data to operate on. In fact, the collections of texts or text corpora are the most valuable assets of text researchers. The text storage and the link to existing digital text archives are therefore a very important part of TextGrid. As long as every archive uses its own retrieval interface, has its own metadata structure and tags its content with a proprietary markup schema, research that combines data from many archives becomes intractable. TextGrid wants to establish itself as a platform that provides on one hand a unified view on the text resources available in TextGrid proper, in archives that are linked with TextGrid or from potential further (commercial) content providers like publishing houses. We are well aware that we cannot force the TextGrid metadata structure, our markup schemata etc. onto existing archives, of course. But we will start with a minimal, DublinCore-inspired subset of metadata that presumably can be mapped to the metadata structure of the relevant archives. If the metadata subset that can be mapped between TextGrid and the external

provider is larger, more complex retrieval queries from within TextGrid can be handled. On the other hand, TextGrid will encourage researchers to publish their results — digital critical editions, lexicons, etc. — including their intermediate findings within TextGrid, so that the next researcher can build upon this data and does not need to repeat all the tedious and very time consuming tasks of adding annotations, fixing markup the automated tools got wrong and so on.

This is facilitated by the TextGrid architecture: In its current revision, it has four layers: The storage elements, the grid middleware, the service layer, and last but not least a (graphical) user interface.

At the storage layer we have any resource that holds texts or metadata. That may be a plain old bunch of disks, relational or XML databases, large file servers, or even whole archives.

It is the responsibility of the grid middleware layer to keep track of where which data is available, maintain a map from logical filenames to physical location, or to decide which data needs to be duplicated somewhere else to increase the overall efficiency of TextGrid. The authentication and authorization infrastructure that allows fine grained access control also belongs in the grid middleware layer as well as the, e.g., the accounting modules.

TextGrid did not have existing ties to any of the Grid middleware systems supported by D-Grid, i.e., to the Globus Toolkit, now in version 4, gLite, or Unicore. From the outset, TextGrid considered building its software on open standards paramount to the projects long-term success and saw the web service technology as well suited to meet its design goals. Therefore, Globus Toolkit 4 (GT4) was chosen as the Grid middleware implementation for TextGrid.

The Globus Toolkit [6] is a framework with many components including execution and data management, securing a grid against unauthorized access of resources, resource monitoring and discovery. GT4 also ships with tools to create and deploy new grid services. Since version 4, many (though not yet all) of the toolkit's services build upon the Web Services Resource Framework (WSRF) [7]. Thus, a Globus grid service is a web service that, in addition to the "usual" web service standards like SOAP [8], WSDL [9], etc., also supports state. The bundle of a web service and the resource that holds the associated state is addressed by so-called *endpoint references* defined by WS-Addressing [10]. The use of established web service technology and, where this is insufficient to meet the complex needs in a grid environment, the implementation of new standards developed by organizations like OASIS and W3C gives confidence that a grid based on GT4 allows for the easy integration of new software and will be interoperable with new technology in the long term.

All the TextGrid tool services run as grid services on top of the middleware. If a service needs, for instance, a particular file, then it contacts some kind of broker service in the middleware layer that determines where this file is stored, selects on behalf of the caller which copy is closest by (for some metric that involves the connection bandwidth and the transmission costs on this line), and finally delivers the file

or the requested parts thereof to the respective tool service. A dedicated workflow tool service allows to use all other services as building blocks to create powerful procedures.

If third parties wish to contribute algorithms to TextGrid, then they need to create and deploy a new tool service. We want to keep the barrier for contributors low; therefore, our intention is to shield the tool services as much as possible from the complexity that is inherent in the management of a grid. For example, all the file handling is delegated to the middleware layer, the tool services may use a very simplified file API. If a service can be implemented as a classical web service, then we do so; this enables the contributors more comfortably to use existing web service libraries or integrate TextGrid services into their own workflow.

Finally, TextGrid will implement an Eclipse Rich Client GUI. For each tool service that can or has to be used interactively, there will be a corresponding Eclipse plugin that provides the actual user interface for the corresponding tool service. In the long term, there may also be a web browser based portal — using, e.g., GridSphere — for tool services that do not require very much interaction with the user.

IV. THE TEXTGRID ECOSYSTEM

A. A Couple of Definitions

When we speak of ecosystems, we follow the general lead of DEST 2007 call for papers and see it as

a self-organizing digital infrastructure, aimed at creating a digital environment for networked organizations (or agents) supporting the cooperation, knowledge sharing and development of open and adaptive technologies and evolutionary domain knowledge rich environments

In our point of view it is crucial, though, to emphasize that these ecosystems are populated by human and digital inhabitants alike and that they are only created, as the call for the special session on eHumanities for Digital Ecosystems puts it, "through the *interactions* between both human and computer-based agents" (our emphasis). They are much more than just "a pervasive 'digital environment' which is populated by 'digital components' which evolve and adapt to local conditions thanks to the re-combination and evolution of its 'digital components'" [11]. Important though the digital infrastructure undoubtedly is, only the satisfaction of and benefit for the various human stakeholders (both institutions and individuals) will make the ecosystem flourish.

B. Stable State

If TextGrid sees itself as a formative project for an ecosystem for the textual sciences and, indeed, of eHumanities in general, then it is worthwhile to imagine how the fully functional system will look like in its stable state. Like all ecosystems, it will involve different species, and like all digital ecosystems those players will be both human and software agents. It will interact and possibly fuse with other, related ecosystems. Historically it has evolved from other states that have left its mark on the collective worldview of its members

and in the future it, in turn, will morph into something else, though exactly what that might be is still pure speculation.

In the following subsections we shall look at these issues in more detail.

C. Previous States

Many of the current players in TextGrid and quite a few of its future (especially institutional) participants have a long involvement in IT-based textual data processing, though not necessarily in collaborative environments or working methods. The concrete software tools that were used vary greatly, though, and many of those have long past their prime, though not without leaving their traces behind even in the new TextGrid ecosystem.

A good case in point may be TUSTEP [12], the Tübingen System of TExtProcessing tools. Originally developed in the late 1960s and continuously enhanced since, it lends itself only with difficulty to a distributed and collaborative working environment and suffers in today's perspective from many shortcomings related to its long and varied history. However, its modular design, its many unique features and its flexibility have formed the expectations of quite a few scholars and will together with other projects such as Tact [13] or Collate (e.g. [14]) help to shape the delineations and the functional requirements for a number of the new software agents — without influencing their actual implementation, though.

D. Subsystems

Like all ecosystems also the TextGrid ecosystem consists of several subsystems that interact with each other to various degrees. Leaving aside for a moment the software agents we consider at least the following three subsystems to be essential for TextGrid's longterm survival:

- 1) End users: the actual scholars — philologists, historians, linguists etc. — that use TextGrid for their research.
- 2) Content providers: institutions — archives, libraries, commercial publishers, but also end users when publishing their materials — that provide quality resources such as images of manuscripts, critical editions, or linguistic corpora.
- 3) Software developers: the team — both computer scientists in companies and academia and domain experts, possibly also dedicated research projects — that maintain the software platform and continue to enhance it.

While all of these subsystems are needed for a thriving TextGrid project, their specific motivations and living conditions are quite different.

The motivation for end users may be most straightforward to pinpoint. They want an environment that enables them to mine existing textual or multimedia resources for their research and software agents that can support their steps towards their final goal, be that a critical edition, a dictionary, a linguistic analysis or otherwise.

Software developers and especially commercial partners have a different goal — especially the latter need to draw tangible benefit (not necessarily in the form of immediate

financial profit, though) from their contributions also beyond the initial funding period (which for the commercial partners is a co-funding model only). As with any open source activity, this can be challenging. Their business plans accordingly point out the service models that will make it in their best interest to continue contributing to the project.

Content providers, especially publishing houses, but also end users or archives will only be prepared to share their data if they can retain full control of their intellectual property. To allow that we will eventually need a digital rights management (DRM) transparently integrated into the technical infrastructure — something that Grids do not inherently support. This requirement was deferred in the original TextGrid proposal as not to make the very first TextGrid architecture too complex and because the content provided by TextGrid members at this stage is free of intellectual property claims by third parties. A new proposal on Service Level Agreements (SLAs) and DRMs is under way, though. These additions will also allow for the inclusion of non-free software agents into the ecosystem that can then be charged for following a variety of payment models. Nevertheless, TextGrid promotes open content and resource sharing.

E. Room for All Sizes

The individual players in all subsystems can differ drastically in size, manpower and technical prowess. Users can and will still often be individuals — the lone scholar will become less lonely, but will not disappear. They can, however, equally be large institutions such as university departments or established research projects spanning many organizations with anything in between.

Similarly, archives can be essentially one or two person activities, but they can just as well be national institutions with a three-figure staff. The same holds true for many of the other players, be they publishers or software companies. One of the key challenges of the TextGrid future will be to establish an ecosystem with niches for all these sizes.

F. Species

For few of the players in the subsystems working with and in TextGrid will be their primary, let alone only concern — not quite unlike biological species that can often live in several ecosystems and cannot afford to be too adapted to any single of them. However, again their specific needs differ. End users and especially individual scholars are often not very technology savvy and need an intuitive user interface on top of the software agents and their interactions. This is not much of a concern for software developers. They, however, in a highly distributed environment will usually have varying skill levels and degrees of familiarity with the system. They profit from clear interfaces, good technical documentation and structured coding guidelines.

External content providers again are often offering their resources in several networks at once and can ill afford significant overhead for each separate integration project. They profit from standardized technologies and standardized interfaces.

To stay in the ecosystem metaphor, the TextGrid Ecosystem must offer its inhabiting species good living conditions and at the same time remain open to its surrounding environment. Individual species — and that includes, as we will see presently, the digital inhabitants — must be able to thrive in it, but they must also be able to migrate to and fro.

G. Interaction with other Ecosystems in Textual Criticism

TextGrid may be currently unique in textual scholarship in that it builds on the grid paradigm and thus on a platform that very much embodies the idea of building virtual organizations (VOs) and digital ecosystems. ([15] [16] [17] [18] are a few examples that discuss the affinity of Grid and VOs / ecosystems.) TextGrid does not exist in isolation, however, nor is it the first project in textual scholarship to embark on heavily collaborative working methods, one of the early ones being the Suda Online (SOL) [19] that was started in 1998.

A few notable new open source software projects in textual scholarship, linguistics, and related disciplines such as cultural heritage have been developed over the last decade and / or are in the process of being developed. These include TAPoR (Text Analysis Portal for Research) [20], Bricks (Building Resources for Integrated Cultural Knowledge Services) [21] [22], Gate (General Architecture for Text Engineering) [23] [24] [25] and the ARCHway (Architecture for Research in Computing for the Humanities through collaborative research, teaching, and learning) Project [26] [27] and its successor EPPT (Edition Production & Presentation Technology) [28].

Some of these projects, especially TAPoR and Bricks, are consciously designed as distributed projects and build on web service standards that Globus Toolkit 4 uses as well (SOAP, WSDL, ...). It is likely that this will result at least in some shared agents that can be called from all of these environments. In the long run they may even merge into a single composite digital ecosystem.

Others, notably ARCHway / EPPT, are less geared towards inter-institutional collaboration (even though EPPT encourages joint editing through version control systems), but as TextGrid sport an Eclipse-based GUI and, e. g. in the case of a link editor to connect manuscript images and transcription, need very similar agents. Also in this case we expect that some of these software agents will eventually thrive in identical or modified form in many ecosystems, adapting in each case to the requirements of the specific environment and typical process chains.

TextGrid aims for collaboration with related projects on various levels. D-Grid demonstrates on the infrastructure level that valuable synergies can emerge from the cooperation of very different communities. From the very beginning, TextGrid got in touch with various activities worldwide that form part of the eHumanities ecosystem. eSciDoc [29] is one of the exciting initiatives that offers the opportunity to collaborate on multiple levels.

H. Communication / Interoperability

As seen before, communication between the TextGrid and other (not only eHumanities) ecosystems is essential. We have

already looked at the case for human inhabitants. For the digital inhabitants of eHumanities ecosystems — or indeed any software systems — to interoperate we need standards that all players adhere to.

Interoperability works on various levels that following the European Interoperability Framework (EIF) [30] can be categorized into three interoperability areas: Organizational, semantic and technical interoperability.

Organizational interoperability is concerned with “defining business goals, modelling business processes and bringing about [...] collaboration” (p. 15). We shall elaborate on this in the next section.

Semantic interoperability “is concerned with ensuring that the precise meaning of exchanged information is understandable by any other application that was not initially developed for this purpose”. TextGrid needs to leverage existing standards for semantic descriptions such as Topic Maps [31] or the Web Ontology Language OWL [32] for the formalization of semantic information. It develops special ontologies for the individual semantic units, be they content or software agents, that can be used in the full, Dublin Core compliant metadata. The metadata entries can be federated and thus made accessible through registries such as [33] and [34].

In the intersection area between semantic and technical interoperability are the content encoding standards that are central to the complete project. Textual content can belong to a number of different text types — novels, dictionaries, drama, poetry, etc. — and within that to different categories — critical editions, linguistic analyses, etc. The guidelines of the TextEncoding Initiative (TEI) [5], currently under revision for the fifth edition, offer general standards for the encoding of a wide variety of text types, but they are often not specific enough for the concrete requirements of some of TextGrid’s software agents, notably the publishing, search and indexing tools. TEI profiles will resolve this problem.

As with most XML-based text encoding standards, TextGrid will use the Universal Character Set (UCS) aka Unicode as its character encoding scheme of choice. In contrast to most, as of yet unencoded special characters / glyphs may be an issue for some editions and are being dealt with in related activities.

Technical interoperability in the strict sense — the “issues of linking computer systems and services” — are covered by a few key standards, both formal and de facto, to which TextGrid and indeed many other projects adhere. These include the use of the accepted stack of Web Service Standards, notably SOAP and the Web Service Description Language WSDL for the communication of software agents, even though the Grid specific Web Service Resource Framework (WSRF) offer some additional benefits beyond a pure Web Service layer that might be needed for some advanced functionality.

On the GUI level, TextGrid opts for the Eclipse framework that is also used in projects such as ARCHway / EPPT and that has developed itself into a kind of de facto standards for cross-platform interactive user interfaces. It should be possible to exchange Eclipse plugins across project boundaries.

V. LIVING IN THE TEXTGRID ECOSYSTEM

We have now seen how the TextGrid ecosystem will look like in a few years from now. In its entirety it is an experiment in organizational interoperability on a large scale — with the advantage that a good part of the definition of common business goals, processes and forms of collaboration amongst the original partners was already done during the application phase. We will need flexibility and maybe even a bit of good fortune to extend this common spirit also to new participants — but we are confident that our approach will scale.

For TextGrid's inhabitants it will change the way they engage in textual scholarship and the manner they conduct their research. Some of these changes are intentional: TextGrid wants to further a culture of collaboration and resource sharing, while maintaining the traditionally high standards of textual criticism and linguistic research. It should become a habit to publish intermediate and final results while leaving the control over what, at which time and for whom it is published with the scholar. Furthermore, the software agents facilitate quantitative and comparative studies across corpora on a scale that might otherwise have been impossible to achieve.

Other changes may be less intentional. As Marshal McLuhan's adage goes, "the medium is the message" [35] — or, more formalized and set into context, no sign process, when mediated through an ICT system, will remain quite unchanged. It is difficult, though, to predict what these side effects will be exactly and to what degree they will be harmful or beneficial.

One danger might be that easy availability of quality resources and software agents could lead to sloppy research, since what is too readily at hand may tempt to do so. We assume that classical peer review will take care of that.

More insidious is the peril that, to have certain software agents at one's disposition may lead to a penchant for research that can easily be done with them to the detriment of other, possibly equally or even more interesting questions that the software agents by themselves or in combination were not geared to elucidate. Very similar is the situation for available content which also may impact the research tasks that a scholar chooses to address. In other words, there is a real risk that tools and existing content may determine the results.

The inherent openness of TextGrid will to some degree counterbalance these dangers — the inhabitants will be able to add new agents and content resources. Then, of course, all technologies including the traditional print approach have other, often equally insidious epistemological curtailments.

On the positive, the benefits of a new climate in textual scholarship will be significant — we dream of an eHumanities ecosystem without restricting boundaries in which research thrives and results flows freely. TextGrid may contribute substantially towards making this become a reality.

REFERENCES

- [1] G. Schmitz and C. Radl, "Die Edition der Kapitulariensammlung des Benedictus Levita," <http://www.zdv.uni-tuebingen.de/tustep/prot/prot841-benlev.html>, Feb 2002.
- [2] Arts and Humanities Data Service, "E-science scoping study," Aug 2006.
- [3] "TextGrid homepage," <http://www.textgrid.de/>, Nov 2006.
- [4] "D-Grid homepage," <https://www.d-grid.de/>, Nov 2006.
- [5] C. M. Sperberg-McQueen and L. Burnard, Eds., *Guidelines for Text Encoding and Interchange*. Humanities Computing Unit, University of Oxford, 2002.
- [6] I. Foster, "Globus Toolkit version 4: Software for service-oriented systems," in *IFIP International Conference on Network and Parallel Computing*, ser. LNCS, vol. 3779. Springer-Verlag, 2006, pp. 2–13.
- [7] OASIS, "Web services resource 1.2," OASIS, OASIS standard, 4 2006.
- [8] W3C, "Soap version 1.2," W3C, W3C Candidate Recommendation, Jun 2003.
- [9] —, "Web Services Description Language (WSDL) 1.1," W3C, W3C note, Mar 2001.
- [10] W3C, "Web service addressing 1.0 – core," W3C, W3C Recommendation, May 2006.
- [11] European Commission, "What is an European digital ecosystem?" <http://ec.europa.eu/enterprise/ict/conferences/doc/p5.de2.pdf>, 2005 2005.
- [12] "TUSTEP homepage," <http://www.zdv.uni-tuebingen.de/tustep/index.html>, Apr 2006.
- [13] I. Lancashire, J. Bradley, W. McCarty, M. Stairs, and T. R. Wooldridge, *Using TACT with Electronic Texts*. MLA, 1996.
- [14] P. M. W. Robinson, "The collation and textual criticism of icelandic manuscripts (1): Collation," *LLC*, vol. 4, no. 2, pp. 99–105, 1989.
- [15] D. D. Roure, J. Frey, D. Michaelides, and K. Page, "The collaborative semantic grid," *cts*, vol. 0, pp. 411–418, 2006.
- [16] Globus Alliance, "An "ecosystem" of grid components," http://www.globus.org/grid_software/ecology.php, Nov 2006.
- [17] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid: Enabling scalable virtual organizations," *International J. Supercomputer Applications*, vol. 15, no. 3, 2001.
- [18] I. Foster, "Globus toolkit version 4: Software for service-oriented systems," in *NPC 2005*, ser. LNCS, H. Jin, D. Reed, and W. Jiang, Eds., vol. 3779, 2005, pp. 2–13.
- [19] R. A. Finkel, R. Scaife, and H.-E. Ng, "The suda project: Collaborative web-based translation," in *HICSS '99: Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 1*. Washington, DC, USA: IEEE Computer Society, 1999, p. 1054.
- [20] TAPoR, "Tapor: text analysis portal for research," <http://tapor.humanities.mcmaster.ca/home.html>, Nov 2006.
- [21] T. Risse, P. Knežević, C. Meghini, R. Hecht, and F. Basile, "The BRICKS infrastructure — and overview," in *8th International Conference EVA 2005 Moscow*. EVA, 2005.
- [22] BRICKS Project, "Bricks — building resources for integrated cultural knowledge services," <http://www.brickcommunity.org/>, Nov 2006.
- [23] H. Cunningham, Y. Wilks, and R. J. Gaizauskas, "Gate: a general architecture for text engineering," in *Proceedings of the 16th conference on Computational linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 1996, pp. 1057–1060.
- [24] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "Gate: an architecture for development of robust hlt applications," in *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 2001, pp. 168–175.
- [25] GATE Project, "GATE, a general architecture for text engineering," <http://gate.ac.uk/>, Nov 2006.
- [26] K. Kiernan, J. W. Jaromczyk, A. Dekhtyar, K. H. Dorothy Carr Porter, S. Bodapati, and I. E. Iacob, "The ARCHway project: Architecture for research in computing for humanities through research, teaching, and learning," *LLC*, vol. 20, pp. 69–88, 2005.
- [27] "The ARCHway Project," <http://beowulf.engl.uky.edu/~kiernan/ARCHway/entrance.htm>, Nov 2006.
- [28] EPPT, "Eppt," <http://beowulf.engl.uky.edu/eppt-trial/EPPT-TrialProjects.htm>, Nov 2006.
- [29] "eSciDoc project," <http://www.escidoc-project.de/>, Nov 2006.
- [30] IDABC, *European Interoperability Framework for Pan-European eGovernment Services*. Luxembourg: European Communities, 2004.
- [31] ISO/IEC 13250, "Information technology — sgml applications — topic maps," ISO, ISO standard, 2002.
- [32] W3C, "OWL Web Ontology Language," W3C, W3C Recommendation, 2 2004.
- [33] OASIS, "ebxml registry services and protocols version 3.0," OASIS, OASIS standard, 5 2005.
- [34] —, "ebxml registry information model version 3.0," OASIS, OASIS standard, 5 2005.
- [35] M. McLuhan, *Understanding media*. Routledge, 1964.