eHUMANITIES DESKTOP
An Architecture for flexible Annotation in Iconographic Research

Rüdiger Gleim, Paul Warner

Dep. for Computing in the Humanities, Goethe University, Georg-Voigt Straße 4, Frankfurt, Germany
gleim@em.uni-frankfurt.de, p.warner@em.uni-frankfurt.de

Alexander Mehler

Texttechnology/Applied Computer Science, Bielefeld University, Universitätsstraße 25, Bielefeld, Germany
alexander.mehler@uni-bielefeld.de

Abstract: This article addresses challenges in maintaining and annotating image resources in the field of iconographic research. We focus on the task of bringing together generic and extensible techniques for resource and annotation management with the highly specific demands in this area of research. Special emphasis is put on the interrelation of images, image segments and textual contents. In addition, we describe the architecture, data model and user interface of the open annotation system used in the image database application that is a part of the eHumanities Desktop.

1 INTRODUCTION

Digital image archives play an important role as a teaching aid in classrooms in many fields, including biology, geography and chemistry. But image archives also serve scientific research as well, for example in the arts and in iconographic research in the historical sciences. When speaking of a certain *digital image archive* the term is usually understood as a combination of the images themselves and the software system delivering them. This tight link is because the software systems are often specially designed for a specific purpose or are an adaption of a more general system for a specific purpose. In the following we therefore distinguish between the *Image Database* (Image DB) as a structured collection of images and the *Image Database Management System* (Image DBMS) as the software system managing the resources. Both Image DB and Image DBMS constitute an *Image Database System* (Image DBS). We focus on systems which are accessible online rather than on offline variants.

Since image database systems are often developed or deployed with a specific purpose in mind there are a vast number of applications which can differ widely in terms of functionality. Commonly this includes a basic means to browse and explore images. Usually some core meta data like name, artist or author, copy-right information and keywords are included. In addition, open systems allow for uploading new images and support *tagging* images with keywords. A popular example is *Flickr*[1]. There are also a number of scientific systems like *Medical Picture*[2] or the *Biology Image Library*[3] which fall under that category. In many usage scenarios this functionality is sufficient.

For scientific research however more elaborate systems are required which offer extensive meta data and allow for expressing image interrelations. *Prometheus* (Dieckmann (2008)) is a distributed image archive for art history, archeology and the cultural sciences. The project was developed at the University of Cologne starting in 2001. Its main target and strength is offering integrated access to distributed, heterogenous image databases for research and education. Prometheus offers a modern web-based user interface to organize and search images. *ConedaKOR*[4] is a commercial web-based image database system for archiving, managing and searching image collections and meta data. It puts special emphasis on the annotation of image interrelations and dependencies which can be used for exploring the

---

[1] http://www.flickr.com
[2] http://bilddatenbank.medicalpicture.de
[3] http://www.biologyimagelibrary.com
[4] http://www.coneda.net/kor/overview

database.

Image database systems, especially very large ones, require well-annotated images in order to support search and retrieval, and to avoid "data graves" (i.e. stored information that cannot be accessed or discovered easily). The degree to which meta data can be annotated depends on the underlying annotation scheme. An annotation scheme defines the data fields which can be used to describe a given image. In most usage scenarios the annotation scheme is fixed and chosen to fit a particular demand. In this article we propose an approach which puts emphasis on flexibility regarding the annotation schemes. Instead of promoting a fixed annotation scheme the (authorized) users of the system should be able to create custom schemes. Furthermore the annotation schemes should be extensible or editable as new requirements for annotating images arise. We also focus on resource management. How can the degree of user access to images, collections and annotations be managed?

Section 2 discusses the challenges of annotating images in iconographic research. Section 3 introduces the eHumanities Desktop, a web-based system for resource management and analysis in the humanities. We focus on the annotating mechanisms which the system offers to annotate arbitrary resources and how they can be used to address the challenges introduced in section 2. Section 3.2 presents the current state of the *eHumanities Image Database System* which already implements important parts of the proposed functionalities. Finally we give a conclusion and prospect of future work.

2 CHALLENGES IN RESEARCH ON ICONOGRAPHIC ANNOTATION

From the point of view of linguistic annotation any sign aggregate (e.g., a sentence, a paragraph or a whole text) can be annotated regarding two dimensions:

- We may segment the aggregate and annotate its internal structure along various functional, semantic or purely structure-oriented annotation models. In this case, we deal with *intra-aggregate* relations (e.g., intratextual relations as bridging (Vieira and Poesio, 2000) or meronymic relations that span logical (e.g., document) structures (Power et al., 2003)).

- Alternatively, we may interrelate the aggregate as a whole with other aggregates and their segments.

	intra-aggregate	inter-aggregate
monomodal	intratextual relations; intrapictorial relations	intertextual relations; interpictorial relations
multimodal	intratextual-intrapictorial (text-segment-to-image-segment relations)	text-image relations

Table 1: The combinatorics of mono- and multimodal relations specified by intra- and inter-(sign)aggregate relations.

In this case we deal with *inter-aggregate* relations (e.g., syntagmatic or paradigmatic relations of lexical units (Hjelmslev, 1969; Barthes, 1977) and, of course, intertextual relations that are established between texts (Thibault, 1997)).

The small spectrum of sign relations that is fundamental for the build-up of sign systems (Hjelmslev, 1969) is complemented by the distinction of different modalities of symbolic and iconographic signs (Kress and Leeuwen, 1996). In this sense, we can distinguish, for example, intra-textual relations of a focal text that relate its iconographic segments with its textual segments. Analogously, we may start with an image to distinguish intra-pictorial relations that interrelate textual segments of an image with some of its pictorial segments. Finally, we have to distinguish relations of sign aggregates as a whole that interrelate signs of the same, or different, modality. In this sense, we deal with text-text, text-image or image-image relations (see Table 1 for a summary of the matrix spanned by these options).

From this perspective we can distinguish the following challenges in research on iconographic annotations:

- *Text segmentation: how to automatically segment texts into their constituents as instances of functionally or semantically demarcated, recurrent segment types* (e.g., rhetorical relations, arguments, propositions etc.)?

- *Image segmentation: how to automatically segment images into their constituents as instances of functionally or semantically demarcated, recurrent segment types* (e.g., emblems, logos, and less conventional, but recurrent image components)?

- *Text linkage: how to automatically interlink texts (and their segments) with each other along intra- or intertextual relations?*

- *Image linkage: how to automatically interlink images (and their segments) with each other along intra- or interpictorial relations?*

- *Multimodal linkage: finally, how to interlink texts (and their segments) and images (and their seg-*

ments) with each other along intermodal relations.

Note that we conceive of image segmentation in semantic or functional terms in that the resulting segments are seen to be semantic or functional units. These are signs in terms of semiotics as they recur in different texts or images to contribute to their meaning constitution (Peirce, 1934; Eco, 1986).

To the best of our knowledge, these tasks are still unsolved. That is, we cannot yet automatically segment texts and images, nor can we automatically identify and type their relations. Interestingly, this especially holds for textual units (in spite of recent advancements in text mining (Feldman and Sanger, 2007) and text-technology). Although there are promising approaches to text segmentation (e.g., (Marcu, 2000)), the automatization of this effort *beyond logical or layout structures* is still out of reach (Stede, 2007) (see Teufel and Moens 2002 for a promising approach in this area). However, this is what we need if we want to explore images and their recurrent components by analogy to *semantic* text retrieval. Consequently, we conceive a kind of *image-related usage semantics*, by analogy to a linguistic semantics, in which significant co-occurrences of pictorial elements are explored as a reference point of their paradigmatic relations of mutual substitutability. Such an approach has been implemented successfully on the level of words and multi-word units (Miller and Charles, 1991; Landauer and Dumais, 1997; Rieger, 2001; Heyer et al., 2006). What we plan to do is to transpose this approach onto the level of images and text-image relations. That is, we conceive an image semantics in which pictorial elements serve as types whose syntagmatic contiguity (or neighborhood) associations and whose paradigmatic similarity associations are automatically computed. These are indispensable ingredients of a semantics that can assign a picture as a whole a meaning which, in turn, is the starting point for interlinking this picture with a text that has a similar (paradigmatically associated) or related (syntagmatically associated) meaning.

At the present time, the only way we see to master this challenge is to support the human (i.e., non-automatic) annotation by text-technological means in order to gain test data by which machine learning algorithms can be trained to foster the semi-automatic annotation of these units. On the one hand, this reminds one of the paradigm of human computation (von Ahn, 2006; von Ahn et al., 2006; von Ahn, 2008) and games with the purpose of image segmentation. On the other hand, we depart from this paradigm in that we do not design games, but aim to enable users to annotate any of the relations mentioned so far, whether mono- or multimodal, reported in Table 1. The subsequent sections describe our current status in preparing and supporting these efforts.

3   eHUMANITIES DESKTOP

Even though computer-based methods are well established in many areas of research in the humanities it is still challenging to bring together resources for analysis with elaborate tools for information processing. Often tools are fixed on an input format, are too complex to be used by a broader public, or address a very specific research problem and cannot be extended.

The *eHumanities Desktop* (Gleim et al. (2009)) aims at integrating an elaborate resource management system with easy to use application modules for work directly on the data. Its scope principally targets all fields of research in the humanities including for example linguistics, the social sciences, and the historical sciences. The design principle of the system is to offer the full functionality of both established and cutting edge processing and analysis methods while keeping it usable for a broad public. The eHumanities Desktop is in the line of systems which aim at the integration of resources and methods in different areas. Chiarcos et al. (2008) presented ANNIS, a general framework for integrating annotations from different tools and tagsets. GATE (Cunningham (2002)) is a system for flexible text categorization and engineering. Clarin Váradi et al. (2008) aim at a large-scale European research infrastructure to establish an integrated and interoperable infrastructure of language resources and technologies.

The primary objective during the development of the system was to achieve a light-weight core system which is *flexible* enough to be adapted to virtually any application, yet *powerful* enough to allow for mass data and distributed, concurrent usage. The core functions include user and group management, repository and document handling and finally their interrelations in terms of access permissions. The core layer of the eHumanities Desktop provides a programmatic API which offers means to control the authorities (users and groups) and the resources they work on. It is designed to abstract from internal representation and storage details. All other components of the system are designed around this core API. This allows for a modular design and integration of new features into the system which stem from various collaborations and research projects.

Typical usage scenarios include distributed research groups that work collaboratively on shared re-
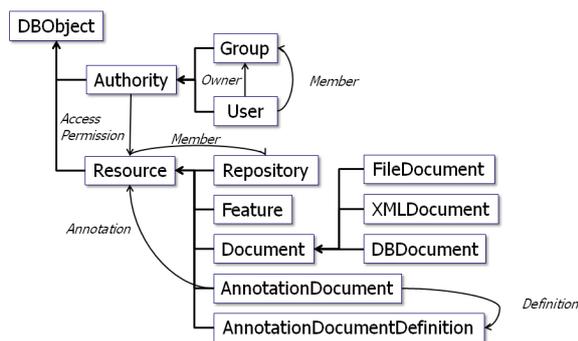
Figure 1: Diagram depicting the master data model of the eHumanities Desktop.

sources. This requires fine-grained access management on both resources and applications. Furthermore, sophisticated means are required to organize documents and repositories. Figure 1 illustrates the data model. The principal concept is that *Authorities* have a certain level of *Access Permission* on *Resources*. The access permission can be either *read, write, delete* or *grant*, as commonly understood from widely-used relational database systems or file systems. An *authority* can either be a user or a group. Users can be members of an arbitrary number of groups and inherit all permissions of those groups. This allows the system to reflect the demands of complex research teams which access and share resources in different roles. Resources on the other hand can be documents, repositories (which help to structure documents), or functions.

3.1 Architecture for Resource Annotation

Elaborate means for resource annotation and queries are crucial for managing and retrieving resources. A major challenge when conceiving an annotation system for a general purpose platform is the wide range of applications. Offering a static annotation scheme like the well known Dublin Core[5] may be convenient for specific solutions but not when the system is supposed to be extensible to new requirements. Section 2 discussed the demands of annotating images for use in iconographic research, which include, for example, typed key-value pairs, hierarchical structures, or typed image interrelations. This section addresses the technical perspective of how a general purpose annotation architecture could be designed to be extensible, yet expressive and performant enough to be applied to specific scenarios like image annotation. The concepts being described are implemented as part of the eHumanities Desktop.

In the following we distinguish between *annotation schemes* defining the data fields which can be used for annotation, and *annotation documents* which are instances of specific schemes. The first question regarding the definition of annotation schemes is what level of expressiveness should be offered. The more expressive the data model, the better the general usability for different purposes. But expressiveness is costly, with a drop in performance when doing retrieval, and it also makes it more difficult to offer the user intuitive means to manage the annotations. A self-evident option would be to rely on *XML Schema* to define the structure of valid annotation documents. XML Schema is a well-established standard and as the annotation documents would be represented in XML they could be processed by a wide range of standard tools. The drawback of that solution is that query performance of existing XML database systems is still not comparable to relational systems- at least when it comes to data oriented structures. Thus internally managing annotations as XML documents would without question be elegant but would slow down performance. Furthermore it would hardly be possible to offer an intuitive user interface for defining annotation schemes to a user who is not accustomed to XML schemas and their complexity.

We decided on a compromise. The user is offered a graphical interface to define typed data fields, can organize them hierarchically into trees if desired and store them as an annotation scheme. Annotation schemes, which define what data fields are available for annotating resources, are represented in XML and based on the *Annotation Document Definition*[6] language. When actually annotating a resource the user first picks which annotation scheme to use and then uses an automatically generated form to fill in the data. Internally both annotation schema and annotation documents are stored in a relational database which offers good performance for retrieval. However it is also possible to export both into XML and process them with other tools.

Now what level of annotation expressiveness is effectively being offered by the system? An annotation schema in the eHumanities Desktop is composed of data fields. A data field has a unique name, description and data type. The data types include strings, numeric and boolean values, dates, and references to other resources. Data fields may be filled with multiple values which can be used for representing keywords for example. The multiplicity can be restricted by specifying the lower and upper bound. In some

---

[5] http://dublincore.org

[6] http://xsd.hucompute.org/add.xsd

cases it may be desirable to define a *value domain* for a data field. A value domain can be specified by enumerating the possible values. Data fields can be structured and organized into trees. Note that the possibility for multiple values for a given data field also includes root nodes of a data field tree. In that case an annotation document can contain multiple independent instances of that data field tree.
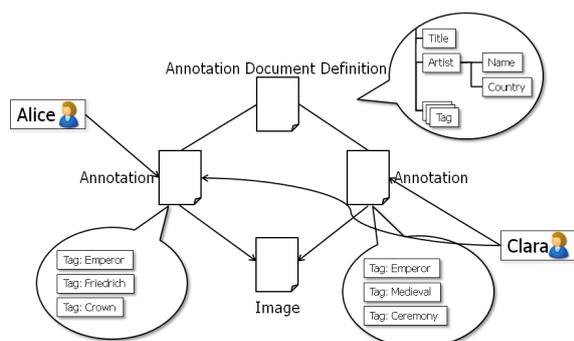

Figure 2: Example of collaborative image annotation.

The master data model of the eHumanities Desktop (see figure 1) regards both annotation schemes and annotation documents as *resources*. As a consequence they are also subject to access management. Thus it is possible to control who is able to access annotations and who is allowed to edit them. Furthermore the data model allows for multiple annotation of a given resource. That way it is generally possible for each user to create a separate annotation document. This approach generally allows for *crowd sourcing* image annotations (for example in classroom) and later on, users could combine or rate the annotations using measures like inter-annotator agreement for example. Figure 2 shows an example of an image being annotated independently by two different users using the same annotation document definition. In this example the user 'Alice' has also given 'Clara' access permissions to her annotation.

Section 2 has put emphasis on the annotation of image sections for subsequent analysis. How could this be implemented using the annotation mechanisms of the eHumanities Desktop? An image may contain an arbitrary number of sections to be annotated. This would be mapped onto a data field which can occur multiple times and serve as a root node for all data fields which are specific for a section. Possible data fields for a section may include the coordinates defining the section as well as keywords and other fields to describe the content. Another approach would be to regard image sections as independent (sub)images which can be annotated separately. The *sub-image* relation would then be represented as a typed reference

between the two images which can also be realized using the annotation system.

In order to search and retrieve annotated documents the user is offered a data grid in which he can choose how to organize, sort and filter resources. An application which offers managing images and annotations as part of the eHumanities Desktop is presented in the next section.

3.2 Image Database System

The *eHumanities Image Database System* (or Image DB for short) is a means for organizing, viewing, and annotating images. As a part of the eHumanities Desktop, the Image DB benefits from a number of other modules in the Desktop, such as the system, user, and permissions control modules, as well as the independent corpus management and annotation modules. Since the eHumanities Desktop is structured in this way, with independent modules, these features are then available to other modules as well. The user and permissions control systems in particular make the Image DB ideal for use by groups of collaborators, or in a classroom. Currently, the Image DB is used for research in the field of historical sciences and as a teaching aid in a university setting.

Images are organized in collections, which are similar to computer file system directories. This structure allows for a simple, easy-to-use application interface, consisting of a window with three tabs: start, collections, and images. The start page, viewed by default when starting the application, displays system or group messages, a list of featured collections, and a display of featured images for the user. These messages and lists can, in the case of groups, be set by the administrator or instructor. There is also a list of recently-viewed collections, so the user can return easily to ongoing projects. Double-clicking one of the thumbnail featured images brings up a larger view of the image in a popup window for better study. The popup can be resized to enlarge or shrink the image.

Clicking the collections tab opens a window in two columns. On the left is a list of collections the user has permission to view, in multiple columns, including name, description and any number of annotation attributes. The collections can be sorted or filtered, or the attributes displayed/hidden, by clicking the attribute column header and accessing a drop-down menu (see figure 3 below). Viewing and filtering by a particular annotation can be done using a drop down menu at the bottom of the collection list, where the available annotations are listed. When a user has many collections, the display is limited by default to 25, and the rest are available via a pag-
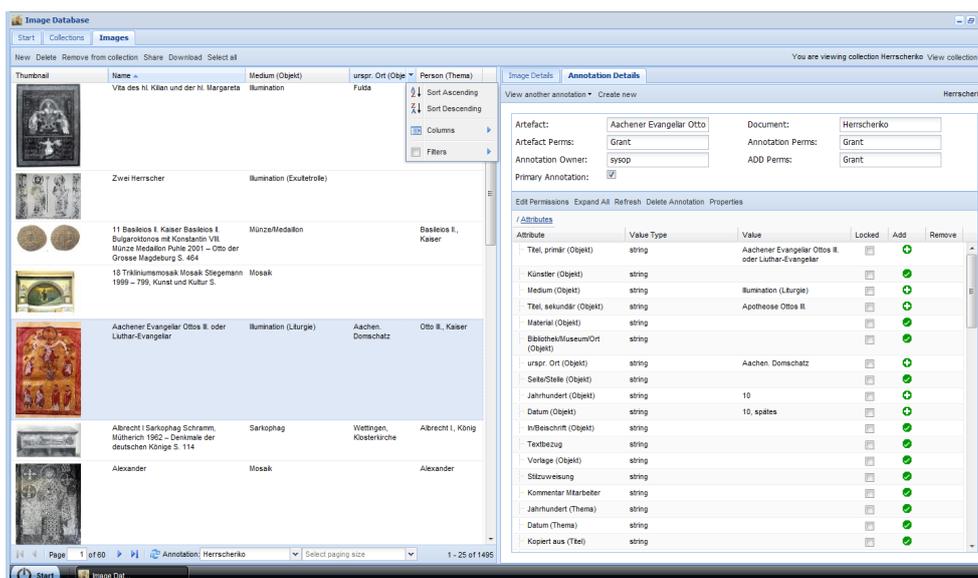
Figure 3: Screenshot of the images panel, showing the annotations tab and the sort/filter menu

ing toolbar at the bottom. The user can quickly cycle through the complete list, viewing 25 at a time, by clicking the next or previous buttons on the toolbar. It is also possible, via another drop down menu in the pager, for the user to define the number of collections that are displayed at one time. Menu items, such as 'new', 'delete', and 'share' are available in a menu bar at the top and via a context menu (opened by right-clicking the mouse). Since permissions are graded in the typical way (read, write, delete, and grant), what a user sees or can edit is subject to fine-grained control. A user with grant privileges can permit other users to see or edit collections or images, including selecting some (and not necessarily all) images within collections for sharing. Annotations are also subject to permissions, and can be shared with, or hidden from, other users.

In the right column, the user can view or edit information or annotations for a selected collection. Certain attributes, such as name, description, creation date, and owner, are available for all collections or images, regardless of annotation. There are two tabs in this column, 'Collection Details' and 'Annotations', for ease of viewing and editing. Selecting a collection in the left column causes its details to appear in the right, so one can quickly scan through the collection list viewing details or annotations, as desired. With adequate permissions, the user can edit attributes in these windows, as well.

Annotations are the means of applying searchable meta-data to the collections (or images), defining them in user-defined terms or according to pre-

existing definitions. Annotations, as noted above, are handled by an independent module in the eHumanities Desktop, and are therefore available to any type of document collection. Annotations are defined by an annotation definition document (ADD), and multiple annotations, using one or more ADD's, can be applied to a collection or image. This way collaborators or students can view the 'shared' annotations of others as well as their own, and use them for searching and analyzing collections or images.

Searching, or filtering, is available via filters in the columns in the collection list on the left side of the page. As mentioned above, one can search by annotation attributes by selecting the desired annotation from the drop down menu at the bottom of the list. This then makes the attributes of the annotation available in the attribute columns. Not all are displayed by default, since some annotations have many attributes; however, the user can simply select which columns to display or hide, and use those for searching/filtering. Filters in multiple columns can be combined, and are applied in the order selected, to quickly provide a result set for analysis. If the user selects 'none' in the annotations drop down menu, then only the general attributes, such as name, description, owner, etc., are available for display, sorting, or filtering.

The interface for the final tab, images, has the same structure and facilities as the collections tab. Everything the user learns and becomes accustomed to in the interface is then transferable between images and collections. Images of course have the additional component that they can be viewed and studied, and

they appear in the right side column as one of the 'attributes' so to speak. Double-clicking the image opens a popup window, as in the selected images on the start page, so the user can adjust the image size for analysis. Images that are too large for viewing in the browser window are automatically resized to fit the available window.

Other important facilities include easy creation of collections with files uploaded from the user's computer, or with selected files from other collections. It is also a simple matter to add uploaded or transferred images to a collection, to remove images from collections (without deleting them from other collections), or to delete images entirely from the database. A user can also easily download images to his or her computer, for analysis or advanced editing.

The eHumanities Desktop, within which the Image DB 'lives', so to speak, is currently used by 110 users organized into 9 groups. The focus of the application lies in linguistic applications like PoS Tagging, lexical chaining and text classification on the one hand and iconographic research on the other. Users include researchers (often working in groups), students and classroom teachers. The system currently manages about 10,000 documents of which about 1,700 are fully annotated images. The integration of a larger image collection of about 50,000 annotated images is planned in near future.

4 CONCLUSION

This article discussed the challenges of annotating images in the field of iconographic research and how such requirements could be met using the annotation system of the eHumanities Desktop. Furthermore the *Image DB* has been presented as a snapshot of ongoing work which already implements a good part of the requirements. Finally we provided information about how the system is currently used by researchers and students. Future work will address the implementation of the graphical user interface for image section annotation and the development of positional queries.

REFERENCES

Barthes, R. (1977). *Image, Music, Text*. The Noonday Press, New York.

Chiarcos, C., Dipper, S., Götze, M., Leser, U., Lüdeling, A., Ritz, J., and Stede, M. (2008). A flexible framework for integrating annotations from different tools and tagsets. *Traitement Automatique des Langues*, 49(2):271–293.

Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computing and the Humanities*, 36:223–254.

Dieckmann, L. (2008). Prometheus - the distributed digital image archive for research and education goes international. In Dunn, S., Keene, S., Mallen, G., and Bowen, J., editors, *Proceedings of the EVA Conference*, pages 61–67, London, UK.

Eco, U. (1986). *Semiotics and the Philosophy of Language (Advances in Semiotics)*. Indiana University Press, Bloomington.

Feldman, R. and Sanger, J. (2007). *The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, Cambridge.

Gleim, R., Mehler, A., Waltinger, U., and Menke, P. (2009). ehumanities desktop - an extensible system for corpus management and analysis. In *Proceedings of the Corpus Linguistics Conference*, Liverpool, UK.

Heyer, G., Quasthoff, U., and Wittig, T. (2006). *Text Mining: Wissensrohstoff Text*. W3L, Herdecke.

Hjelmslev, L. (1969). *Prolegomena to a Theory of Language*. University of Wisconsin Press, Madison.

Kress, G. and Leeuwen, T. v. (1996). *Reading images: the grammar of visual design*. Routledge, London.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.

Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, Massachusetts.

Miller, G. A. and Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.

Peirce, C. S. (1934). *Pragmatism and Pragmaticism*, volume V of Collected Papers of Charles Sanders Peirce. Harvard University Press, Cambridge.

Power, R., Scott, D., and Bouayad-Agha, N. (2003). Document structure. *Computational Linguistics*, 29(2):211–260.

Rieger, B. B. (2001). Computing granular word meanings. A fuzzy linguistic approach in computational semiotics. In Wang, P., editor, *Computing with Words*, pages 147–208. Wiley, New York.

Stede, M. (2007). *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik*. Narr, Tübingen.

Teufel, S. and Moens, M. (2002). Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

Thibault, P. J. (1997). Intertextuality. In Asher, R. E., editor, *The encyclopedia of language and linguistics*, pages 1751–1754. Pergamon Press, Oxford.

Vieira, R. and Poesio, M. (2000). An empirically-based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–593.

von Ahn, L. (2006). Games with a purpose. *IEEE Computer Magazine*, 39(6):92–94.

von Ahn, L. (2008). Human computation. In *IEEE 24th International Conference on Data Engineering (ICDE 2008)*, pages 1–2.

von Ahn, L., Liu, R., and Blum, M. (2006). Peekaboom: a game for locating objects in images. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 55–64, New York. ACM Press.

Váradi, T., Krauwer, S., Wittenburg, P., Wynne, M., and Koskenniemi, K. (2008). Clarin: Common language resources and technology infrastructure. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., and Tapias, D., editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).