Institut für Informatik

Programmier– und Modellierungssprachen

Lehr– und Forschungseinheit 4

Prof. Dr. François Bry

Ludwig Maximilians Universität München

The ConStruct Project

(Content Structuring of Raw Texts)

Robert Reiner

Thomas Traber

Abstract

Information on the Internet is often available only in the form of raw texts. Since there are only a few possibilities to format a raw text, the formatting of these texts is often ambiguous. To process the texts (e.g. in an information extraction system) it is often useful to have the text information separated from the formatting. For some applications (e.g. information retrieval) it is even more useful to have the structure of the document available as well.

This paper describes a system that tries to recover the structure of such documents as far as it is expressed by their poor formatting. To achieve this goal, we used the blackboard system architecture with experts for the various different structures of texts.

Fortgeschrittenenpraktikum

Beginn der Arbeit: 01.08.1996
Abgabe der Arbeit: 23.04.1997
Betreuer: Prof. Dr. François Bry
Dr. Norbert Eisinger
Dipl.–Inform. Tim Geisler

1. Introduction

Information on the Internet is often available only in the form of raw texts. Since there are only a few possibilities to format a raw text, the formatting of these texts is often ambiguous. To process the texts (e.g. in an information extraction system) it is often useful to have the text information separated from the formatting. For some applications (e.g. information retrieval) it is even more useful to have the structure of the document available as well.

This paper describes a system that tries to recover the structure of such documents as far as it is expressed by their poor formatting. To achieve this goal, we used the blackboard system architecture with experts for the various different structures of texts.

ConStruct is a team project of two members, done as a *Fortgeschrittenen–Praktikum* at the Ludwig–Maximilians–Universität, München. The subject of the work originated from ourselves as part of another project, whose purpose is to extract information from text documents (mainly eMail and news). ConStruct is therefore the first step to this goal since it prepares the text input for further processings.

The practical work was done in approximately three months. The documentation took another two months.

This report is divided into nine chapters: First of all we will describe the problem and the main goal of this work. The third chapter presents SGML (*Standard Generalized Markup Language*) which is used to represent the structure of the examined texts. The following four chapters will introduce the concept of the blackboard architecture and the way we implemented the blackboard, the control and the knowledge sources. After this we give an overall view on how the system works. In the final chapter, we summarize our results.

2. The Problem

Many of our daily used documents, such as eMail messages or news articles, are plain texts. These texts are written using the ASCII or – still infrequently – one of the ISO–8859 encodings.

ASCII (American Standard Code for Information Interchange) is a 7–bit character set and the ten different ISO encodings are all 8–bit character sets. Even if today's hardware and software is capable to work with 8–bit texts, it is common to use the more widely disseminated ASCII for texts like eMails and news articles. Within the ASCII character set there are 95 printable characters and some control characters (including the carriage return and the tabulator). The ISO 8859 standard includes several 8–bit extensions to the ASCII character set. ISO 8859–1, the "Latin Alphabet No. 1", has become widely implemented and may already be seen as the de facto standard ASCII replacement.

Plain texts are written and read using the encodings described above and displayed using a non–proportional font (a font in which all characters have the same width). Since all characters (except the tabulator) have the same size, the text as an entirety can be seen as a two dimensional matrix of characters. There is no way to change the style or size of the font, to highlight some words by using a different slant (e.g. italic) or weight (e.g. bold). This is the reason why it is nearly impossible to format a raw text properly. However, the simplicity of raw texts made it easier to build platform independent message systems like the Internet eMail and Usenet news system – even if the possibilities of modern computer hard– and software will not be exploited.

Usually the format of the documents we read allows us to make out the structure of the underlying text easily. Even the computer needs an explicit marked up structure of the text to process it (e.g. in information retrieval or information extraction systems). To get around these problems there are two different solutions:

- The eMail and news software systems should be able to handle other texts than raw ones. Various companies (e.g. Netscape Communications Corporation) work on implementations of this solution. It remains to be seen which of these attempts carries through.
- Auxiliary systems can try to extract the structure out of the poorly formatted raw texts by examining the format of the texts.

This project is an approach to the second solution.

The objective of this work was to specify and develop a system that tries to (re)construct the structure of a raw text out of its (rather poor) format.

There are some difficulties to solve. First, there is almost no convention how to use formatting in raw texts to express a certain structure. Second, there is an imbalance between the number of possibilities to format a raw text and the number of used structures in texts.

The possibilities to format a raw text document are restricted to:

- (Meta–)texts (e.g. "`Subject: `", "`PS: `"),
- indented lines,
- new lines,
- capitalization, and
- special characters that can be used to "draw" (e.g., underlining of some words can be simulated with the character "`=`" in the line below the words, whole tables can be formed using the characters "`+`", "`-`", and "`|`"; in eMails and news articles words are often highlighted by preceding or trailing multiplication signs or underscore characters (e.g. "`This is a very *important* message!`").

These few possibilities contrast with an enormous amount of possible structures in text:

- chapters,
- paragraphs,
- headings,
- quotes,
- lists (sorted, unsorted, ...),
- tables,
- emphasizings,
- ...

This imbalance shows that there cannot be an unambiguous mapping between the possible formattings and the possible structures of texts.

This paper describes a system that tries to recover the structure of plain text documents as far as it is expressed by their poor formatting. We used the blackboard architecture as the basic structure for solving the problems of structure recognition and the Standard Generalized Markup Language (SGML) to represent the structure of the texts. SGML is briefly introduced in the next chapter. The blackboard architecture is well suited for our problem as will be shown in chapter four. Afterwards three chapters deal with the blackboard, the control, and the knowledge sources in more detail. The last chapter introduces the working method of the whole system.

The entire project was implemented using the new programming language Java (Version 1.0.2). Java has been developed by SUN Microsystems, it is object–oriented and architecture neutral [1]. According to Grady Booch [2], the object–oriented design is very well suited to the domain of blackboard systems. In our résumé we will discuss some of the problems that originated from this choice.

3. Using SGML for Structuring Documents

SGML – An ISO standard

The Standard Generalized Markup Language (SGML) is a meta–language by which other markup languages are defined; e.g., HyperText Markup Language (HTML). As a markup language SGML is a language which is used to define the structure of a document by introducing special code (markup) to the text and to deal with the task of document exchange on different computer systems.

Precursor of SGML (Goldfarb et al., 1970 [3]) is IBM's Generalized Markup Language, which was created in conjunction with the Computer Graphics Association in the late 1960s. SGML has – amongst others – been used by the United States Department of Defense and the Office of Official Publications of the European Communities before the standard was formally approved by the Text and Office Systems Subcommittee of the International Standard Organization (ISO) in October 1986 [4].

SGML is very successful.

> *"Within two months of the standard's publication, ISO announced that it was the fastest selling standard they had ever produced, having already out–sold the ten–year–old specification for the FORTRAN programming language" [5]*

Structure, Content, and Format

Central to SGML is the concept that documents consist of three major components: structure, content, and format: `Structure` is the document's logical organization. It defines the order in which the single elements are assembled and the logical relationship between them. Markup is used to indicate the different parts of the document instance. `Content` is the actual data within the document. `Format` is responsible for the presentation of the content and plays a major part to distinguish visually the different structural components of a document. Although format is a major component of documents, it is not considered to be an integral part of SGML.

SGML documents

`SGML documents` have three sections: SGML declaration, document type definition, and document content.

SGML declaration

The `SGML declaration` defines the syntax and character set used within the document. One part of this declaration is the definition of the `concrete syntax`; i.e. the syntax that, for example, describes how a tag looks like.

Since there is often no need to alter the appearance of a tag or the used character set, SGML provides a default concrete syntax called reference concrete syntax.

Document type definition

The document type definition (DTD) contains declarations defining the structure of a certain type of document. This is also called the abstract syntax. For instance there can be a DTD to describe the structure of an eMail. In this DTD the author will determine that an eMail consists of a header followed by a body. The body may be empty. Within the header there can be fields called author, subject, date, etc. These fields can appear in any sequence. Adding some further declarations, the DTD will determine the structure of every document of the class eMail.

We now will introduce two of SGML's thirteen declarations: The element and the entity declaration. These are used frequently in our project.

– Element declaration

  Elements are the major structural components of a document. They give a name to a structural element (e.g. QUOTE, PARAGRAPH) and declare which components can be found within this text object (content–model). A (simplified) element declaration looks like this:
  <!ELEMENT *name* (*content–model*)>
  Introduced by the reference concrete syntax sign for declarations ('<!'), the keyword ELEMENT (also taken from the reference concrete syntax) indicates that this is the declaration of the element with 'name' as generic identifier.

  Example

  <!ELEMENT MAILQUOTE (#PCDATA|PARAGRAPH|MAILQUOTE)+ >
  This declaration introduces the element MAILQUOTE. Within a MAILQUOTE there can be either further MAILQUOTEs or PARAGRAPHs (multiple occurrences indicated by '+'). The reserved word #PCDATA indicates that within a MAILQUOTE there can be 'pure' text. PCDATA stands for Parseable Character Data. This means that the data has to be parsed, because there might be further tags or references contained. The counterpart to PCDATA is CDATA, which represents data within which no further SGML tags or references are resolved. In between these two types of data is RCDATA (Replaceable Character Data). Within this kind of data, SGML references will be substituted.

– Entity Declaration

  Entities introduce a special character, include external information, serve as a shorthand notation, or be used as variables. We give only the syntax to declare a general entity, which is used to define special characters and symbols within a document.
  <!ENTITY NAME "*entity value or content*" >
  In our project we use only this kind of entity. Within the document instance, general entities are introduced by an ampersand ('&') and end with a whitespace character or semicolon (';').

  Example

  It is important to replace all occurrences of the less–than character ('<') by a general entity reference, if the reference concrete syntax is in use. This syntax defines the less–than character to be the introductional delimiter of a tag. The declaration to replace every occurrence of '<' by '<' is given below.
  <!ENTITY lt "<">

Document content

The content is the text of the document, including data characters (providing the information of the document), and markup characters.

Example

The following shows an example of an SGML document instance for an eMail. Note that the entity references ('<' and '>') will be replaced by an SGML parser with '<', and '>' respectively.

```
<EMAIL>
 <HEADER>
   <SUBJECT>Sample text</SUBJECT>
   ...
 </HEADER>
 <BODY>
   ...
   <MAILQUOTE>
     <PARAGRAPH>
       A sample text introducing Thomas' email address:
       <BLOCKQUOTE>
         &lt;traber@informatik.uni-muenchen.de&gt;.
       </BLOCKQUOTE>
     </PARAGRAPH>
   </MAILQUOTE>
 </BODY>
</EMAIL>
```

In this example every content is enclosed by a start tag (e.g. `<BODY>`) and an end tag (e.g. `</BODY>`). In the reference concrete syntax the '<' is called the start tag open delimiter (`STAGO`), '</' is referred to as end tag open delimiter (`ETAGO`). In both cases the end tag close delimiter (`TAGC`) is '>'.

Why we use SGML

The reason why we use SGML for the representation of our data–structure is simply the fact that it is a recognized ISO standard that proved its work–ability over a period of more than thirty years.

This is the reason why there are numerous tools that support SGML documents (browsers, DTD generators, ...).

Furthermore, SGML coded documents are easy to read for a human, since they lack of cryptic structuring instructions.

Implementation

SGML is the basic data structure of the ConStruct project. It consists of three parts: the declaration, the document type definition, and the document instance. Since the DTD is not really needed within our project, it is only rudimentarily implemented.

Declaration class

The Declaration class provides hash tables for delimiters, separators and general entities as given in SGML's reference concrete syntax and implements a simple SGML parser. The parser is based on the rules given in chapter 9 of Goldfarb's SGML Handbook [3]. We implemented only those parts of the parser we needed in our project. Refer to the API for more details.

Document Type Definition class

As stated above the Document Type Definition class (DTD class) is not needed in the ConStruct project. At this stage we provide only a skeleton that could be fleshed out in further development.

Instance class

The instance class is the class that provides methods to store information in the hierarchical data–structure and retrieve information from it. The root of this structure is the base document element, stored in the `DocumentInstance` object.

Each node may be of one of the following classes. Each of these is a subclass of the `Content` class.

`Container`:

A container capable to contain further `Content` objects.

`Data`:

A content that represents some sort of data. At this stage of development only CDATA and PCDATA are implemented.

`GeneralEntity`:

A content that represents general entities.

`Empty`:

A content to represent the empty content. This class is not needed yet, but implemented.

The `Instance` class provides an empty slot in which additional information can be stored. In the ConStruct project we use this slot to hold the coordinates of the original text, contained within the Content object (Container, Data, ...).

Perspectives

We recognized early that we never had the chance to implement all properties of a SGML class in the scope of this project, but we decided to stay as close as we could to the ideas of SGML. We also tried to find solutions that are easily extensible.

The work that could be done here is tremendous. From the SGML parser, completing the reference concrete syntax and the ability to read further concrete syntaxes are probably the next steps to complete the `Declaration` class.

The `DTD` class provides only the idea of how it could look like. Methods to read and write DTDs and to prove the correctness of an SGML document could be next to be implemented.

Extending the ability to use style sheets and improving the access times by buffering results on previous requests are two possibilities to improve the usability of the `Instance` class. Further feeatures could include the ability of writing the data to different formats (LaTeX, ...).

And there are numerous other SGML features waiting to be implemented...

4. Blackboard Systems

The idea

The blackboard system architecture is an approach to coordinate problem–solving. It was first proposed by Newell in 1962 and used for the first time by Erman and Reddy in their speech recognition systems `Hearsay` and `Hearsay II` [7]. Since the blackboard model has proved its usability for the problems of speech recognition, a variety of blackboard–based systems have been developed and used in this and many other domains. Blackboard systems have been applied successfully to interpretation problems such as speech understanding, signal understanding and image understanding, planning and scheduling and arrangement–assembly (structure identification)[8].

The blackboard model is often explained by an analogy to a group of people solving a jigsaw puzzle:

> *"Imagine a room with a large blackboard and around it a group of people each holding over–size jigsaw pieces. We start with volunteers who put on the blackboard (assume it's sticky) their most 'promising' pieces. Each member of the group looks at his pieces and sees if any of them fit into the pieces already on the blackboard. Those with the appropriate pieces go up to the blackboard and update the evolving solution. The new updates cause other pieces to fall into place, and other people go to the blackboard to add their pieces. It does not matter whether one person holds more pieces than another. The whole puzzle can be solved in complete silence; that is, there need be no direct communication among the group. Each person is self–activating, knowing when his pieces will contribute to the solution. No a priori established order exists for people to go up to the blackboard. The apparent cooperative behavior is mediated by the state of the solution on the blackboard. If one watches the task being performed, the solution is built incrementally (one piece at a time) and opportunistically (as an opportunity for adding a piece arises), as opposed to starting, say, systematically from the left top corner and trying each piece" [9].*

Now, we will extend the analogy to introduce the missing, last component of a blackboard system: To allow the group to solve the board even if the solution evolution on the blackboard has lead to a dead end, the players are also allowed to remove pieces from the blackboard. Assume that the given rules lead to an enormous confusion, since people add and remove more and more pieces from the blackboard. An organizer is introduced: His task it is to decide which one of the applicants is allowed to contribute to the solution.

The components

The analogy shows that the blackboard model consists of three components: one blackboard, multiple knowledge sources (the people) and one control system (the organizer).

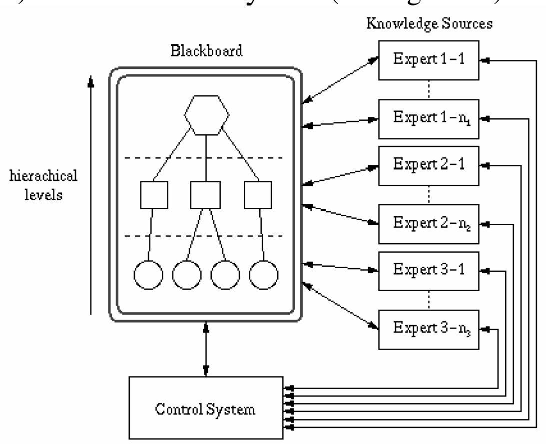

Figure 1: Structure of a BBS

The `blackboard` is a global database that is shared by all experts. It contains the data and (eventually) hypotheses (i.e. potential partial solutions). Nii wrote about the blackboard:

> *"The purpose of the blackboard is to hold computational and solution–state data needed by and produced by the knowledge sources. The blackboard consists of objects from the solution space. The objects on the blackboard are hierarchically organized into levels of analysis. The objects and their properties define the vocabulary of the solution space." [10]*

The problem–solving knowledge is spread over the `knowledge sources` (also known as experts). They can be represented as procedures, sets of rules, or logic assertions. The experts interact via the blackboard, which means that they are independent from each other. They can read the blackboard and even write on it, if they are allowed to by the control.

In order to organize the process of problem solving, all blackboard system models use a `control` system (also known as scheduler) that has to decide which knowledge source should make its contributions to the current blackboard. In some models the control system asks the experts for potential contributions, in some they are self–activating: If certain preconditions are meet, certain actions are applied.

Since the first blackboard systems, many variations of such systems have been developed. The only common thing of these variations is the composition of the three components blackboard, control and knowledge sources. As a result of this diversity there is no standardized terminology.

How it works

The basic cycle in the problem solving process consists of the following steps:

1. The last changes of the current blackboard will be examined, to decide for which of the knowledge sources they are relevant.
2. If there is no relevant expert found any more, the blackboard is solved and the algorithm stops.
3. The control system selects one of the knowledge sources that were determined in the first step. The selection criteria may depend on some rating of the experts and other factors.
4. The chosen knowledge source makes its contributions to the blackboard.
5. The process continues with step 1.

The problem solving by blackboard systems is incremental and opportunistic: The final solution on the blackboard is constructed step by step by the knowledge sources. This is also known as incremental hypothesize and test. Since there is no fixed strategy or algorithm for the solving process, the problem solving is called to be opportunistic. Which action is taken next is determined considering only the current status of the blackboard.

The blackboard model provides great flexibility in structuring problem solving, but naturally this flexibility implies also one disadvantage: Effective control is very difficult to realize, since it is very complicated to estimate the expected value of potential contributions. In this domain much research work is done currently.

Why we use it

We applied the blackboard system architecture to the problem of structuring raw texts. The blackboard contains the text and the recognized structures. There are experts each of which is able to recognize a certain formatting.

The blackboard architecture has been developed to deal with the problems of speech understanding, but it can be applied to many other domains, where the following problem characteristics are given:

- The search space is very large.
- The input data is errorful and/or incomplete.
- There is only an imprecise and/or incomplete problem–solving knowledge. [8]

All the characteristics described above are met by our structuring problem.

Large search space:

Texts are usually full of structures that can be combined in almost any way. Many of them can contain other structures within. This results in an enormous amount of possibilities, which structure to recognize first and which next and so on. Additionally there can be formattings which can be interpreted in different ways. This leads to a very large search space, which can be handled by the blackboard system architecture through its opportunistic problem solving strategy.

Errorful/incomplete input data:

eMails are often of an awfully low quality and News articles even worse: For example, structures like paragraphs or lists are not separated properly, the formatting is not consistent and the spelling is often incorrect. In quite a few texts some of these error appear together. Using a blackboard system the resulting problems can be solved comparatively easy. Every expert is responsible for the recognition of a certain document structure and will rate each occurrence of this structure. If the formatting is messy, the rating will be decreased. The expert’s contribution will only be accepted by the control system, if there is no other expert that assigns a higher rate to the same portion of text. The experts do not have to worry about interactions with other experts.

Imprecise problem–solving knowledge:

The formatting of raw texts is not standardized in any way. Each single author has got his own habits of formatting his texts. Since we only have a very imprecise and incomplete problem–solving knowledge, the blackboard architecture is of great help. The model gives the conceptional frame for the program and allows explorative programming.

Our implementation

Our implementation of the blackboard system consist of one control system (see chapter six, one blackboard (described in the next chapter) and multiple knowledge sources (see chapter seven). The whole system can be composed according to a declarative configuration file or string. For further details on this configuration read chapter eight. The configuration file makes it possible to configure the system without having to recompile the sources.

5. Blackboard

How it works

The `Blackboard` class is the class that stores the document that will be analyzed by the experts. The storage is done by an SGML object. This class provides access to the SGML object.

Its first task is to put the text – given as a string or a file – on the blackboard. After this the content of the board is an SGML document with one UFO Container (UFO – Unidentified Floating Object) as the document base element. The text inside this base element is stored as PCDATA within a `Data` object. The text itself contains no characters that can be confused with SGML delimiters, since all occurrences are replaced with references to SGML general entities. This replacement is done during the initialization process.

On the board each SGML object has one additional attribute representing the coordinates of the original text this object covers. Initially the UFO Container covers the whole text. So its coordinates point from (0, 0) to (lastColumn, lastRow).

As described in the previous chapter, the structure on the blackboard is organized in hierarchical levels. A knowledge source recognizes certain document elements (e.g. a paragraph) and builds one or more new SGML objects with distinct generic identifiers (e.g. "`PARAGRAPH`"). Within this recognized element there may be further structural components of the document. The recognition of these components will be done on the newly built elements of the next hierarchical level (e.g. the `EmphasizeExpert` will only work on UFOs inside paragraphs).

At the end, when no expert has something to contribute, the board is cleared from data not needed anymore. At the current stage of our work this clean–up only erases empty UFO containers. These containers are formed because of some missing experts. Their absence makes it necessary to have the empty lines available up to the end. If the document is fully recognized this information can be discarded.

Perspectives

Improvements that can be done to the `Blackboard` class are mainly the addition of backtracking and the possibility to add additional information to the blackboard that can be used by the experts (e.g. name of the author of an eMail, extracted by the eMail–Expert, can be used by the experts for eMail–signatures).

6. Control System

How it works

We have implemented only one, very simple control system, which uses a straightforward algorithm for solving the blackboard.

In each round of the process of solving the blackboard, each expert is asked in the order of its appearance in the configuration file. After all experts have been queried, the one having returned the highest rating is given access to the blackboard, enabling it to make its contribution. If two or more experts return the same rating, the first expert specified in the configuration file is chosen.

The `Control` class provides two important methods called `nextStep` and `solve`. The method `nextStep` enables the user to solve the blackboard round by round; `solve` solves the blackboard completely.

An exemplary run of the system, illustrating the functionality of the complete blackboard system, will be given in chapter eight.

Perspectives

A lot of efficiency can be gained by improving the control system. One possible amendment could be to divide the rounds in different turns. In each turn only certain groups of experts are asked to rate the board. At first only experts recognizing types of documents could be employed. After one has made its contribution, there would be no need to ask the experts of this class again. Some of the knowledge sources recognize only smaller text segments. These can be asked when all experts who recognize major components have been queried and have nothing to contribute (anymore). Experts that work on pure texts could be asked when all other experts have nothing to contribute. This method would greatly improve the runtime of the blackboard system.

Another improvement could be to give several experts simultaneous access to the blackboard within one round to make their contributions. Naturally this can only be done if the recognized areas do not overlap. Each additional expert that fires during a certain round and that is given access to the board to make its contributions would diminish the total number of rounds by one. Since all knowledge sources have to re–examine the whole blackboard in each round, this modification would mean a substantial reduction of the time required for problem solving.

If there were some knowledge sources which are able to finally rate the current solution (on a certain hierarchical level), then it would be useful to allow the control system to backtrack and use the knowledge sources in an alternative order.

7. Knowledge Sources

The Expert Class

The central method of this class is `ask`. This method will return a list of `RatedCoords`. Note that there can be more than one recognized text area on the blackboard. For each recognized passage an expert will return a corresponding `RatedCoord` object, which contains information about what text area has been recognized by the expert, at what rating, and some additional information.

The `ask` method often analyses the given SGML contents and prepares the modifications to the board without actually committing them. This is done due to efficiency aspects. Analysing is very expensive, while building a new content from the recognized information is comparatively cheap. If the expert is called to make its contributions (method `doIt`) to the board, it has to simply exchange the relevant content(s) with the prepared one(s) without analysing the blackboard again.

The rating is a value between 0.0 and 1.0. The rating 0.0 indicates that the text area in question is clearly not recognized as a text structure the knowledge source is an expert for. 1.0 is a definite sign that the text area is recognized.

All experts of the blackboard system inherit from the base class `Expert`. This class provides two fields. One is the name of the expert, the other is a factor that will be multiplied with the calculated rating. This way it is very easy to calibrate the rating of a special expert, since this factor can be set either in the configuration file or during a single run. The default factor is 1.

The Knowledge Sources in Detail

In the following we introduce our experts that have been developed so far in alphabetical order.

EmphasizeExpert

The `EmphasizeExpert` is an expert for emphasized texts, i.e., text–passages that have been marked up with ASCII characters. A typical form of emphasis would be adding an '*' at each side of the emphasized word (e.g. "`This id-card must *not* be folded!`").

The special characters ("*", "_", "+", and "!") indicate the intended nature of the emphasis. The asterisk indicates bold, the underscore underline. '+' and '!' reflect unspecific emphasis. The generic identifier of the SGML tag will be EMPHASIZED. The style of the emphasis (bold, underlined, unspecific) is indicated by an attribute of the SGML element. The content of this tag will be the text between the two (or more) special characters (e.g. "`This id-card must <EMPHASIZED style="bold">not</EMPHASIZED> be folded.`").

MailExpert

This expert is specialized on the recognition of eMails. The format of eMails is specified by RFC 822 [11] (Request for Comment). All eMails messages are composed of two parts: The header, which contains mandatory (e.g. the sender and addressee of the eMail) and optional information (e.g. the subject of the message) in a fixed format and the body, which can be a freely formatted ASCII text.

Due to the fact that the format of Usenet news articles is also described by this RFC, the `MailExpert` has to care especially about the eMail specifics. Both the `MailExpert` and the `NewsExpert` inherit from a common abstract class `MessageExpert`. This abstract class provides all methods to recognize messages which are RFC 822 conform and evaluates their headers.

NewsExpert

The `NewsExpert` identifies news articles from a Usenet news system. Like eMails the format of these articles is specified by RFC 822 (see above).

ParagraphExpert

The `ParagraphExpert` recognizes paragraphs in documents.

It is important for this expert that every line of the paragraph starts at the same horizontal position. Only the first line may break with this rule and may start at a position prior or after the rest of the lines. At the moment centered paragraphs will not be recognized.

A line exclusively consisting of whitespaces (this includes especially empty lines), marks the end of a paragraph.

The number of lines of the recognized area is important for the final rating. The more lines, the higher the rating. For instance a text area of three lines would give a basic rating of 0.7, four or more lines would yield 1.0.

The expert also recognizes left or block alignment. At the moment right alignment will not be recognized.

Perspectives

Another criterion for recognizing a paragraph could be the approximate equality of the line lengths. However, finding the starting point of the paragraph could be a difficult task.

The `ParagraphExpert` and the `PreformatExpert` are doing the same work twice. To increase the efficiency of the implementation one expert performing the two tasks should be considered.

PreformatExpert

The `PreformatExpert` recognizes preformatted text areas.

To recognize a preformatted text area the expert analyses the occurrences of blocks of specific groups of characters. These are at this stage some special characters (punctuation marks, "#", ...) and whitespaces. The more special characters occur, the higher will be the rating for the analysed text area (e.g., within a table there will be a more–than–normal rate of '–', '+', and '|').

QuoteExpert

In eMails and news articles preceding texts are often quoted to put the new text into the context. These quotes are often introduced by a short text (e.g. "`Karl Lem <lem@domain.org> wrote:`") in which the author of the quoted text and other information are stated. Each line of the quote begins normally with one or more special characters as ">" or ":".

The `QuoteExpert` recognizes the quotes that are formatted as described above. To recognize recursive quotes the expert has to be asked and used repeatedly, since it can only find one level at one time.

Perspectives

In its current implementation this expert does not try to search for the introducing texts, even if they are often created automatically by the eMail or news reader. It would be useful to extract as much information as possible from the introductions of the quotes. This should be possible at least for the automatically created quotes.

SalutExpert

This expert recognizes salutations and signatures dependent on the salutations introductions (like "Dear", "Hello", "Regards", "Bye", ...) provided by the configuration file.

The analysis ends as soon as a line of whitespaces, two special characters ('+', '–', '~', '_', '=', '.', '<', '>') in front of a line or more than a specified number of unrecognized words are encountered. Dependent on what has already been processed and recognized as part of a salutation, the expert returns its rating.

Perspectives

As an improvement for the `SalutExpert`, a database of first names could be provided.

The name of the addressee and the name of the author could be used to improve the possibility to recognize the salutes.

SignatureExpert

The `SignatureExpert` is an expert for signatures typical in eMail and news. Signatures often start with two minus signs. They often include the eMail address of the author and the URL of his homepage. This expert tries to make use of these conventions and of the fact that signatures usually are the last lines of an article or an eMail.

SubDocExpert

This expert demonstrates the modularity of the blackboard system. The expert has its own configuration file upon which it initiates a new ConStruct blackboard system with its own control, blackboard and knowledge sources. In the configuration file only the media experts of the ConStruct system are specified (i.e. the `MailExpert` and the `NewsExpert` at the current stage of development).

With this blackboard system the `SubDocExpert` analyses all UFOs that are contained within a quote. If the auxiliary blackboard system can solve its board, the text within the UFO will be tagged as a sub–document.

TextExpert

This expert recognizes pure text sections in paragraphs, if these sections do not contain any further structural informations. The `ParagraphExpert` builds PARAGRAPH containers, which contain the whole content of the found paragraph within an UFO container. Within this UFO further structures can be found (e.g. lists or emphasises). If there are no structures any more, there is no need for the UFO containers.

Our implementation of this expert rates any UFO container within a PARAGRAPH container with a very low value. This allows the control to let all the other experts contribute there solutions before the `TextExpert` does its work.

Perspectives for all Experts

Since implementing knowledge sources in Java is really a hard job, it would be very useful to have another abstraction layer: To implement the experts in a declarative way a new programming language would be needed. This language has to provide the data structures we used (SGML) and a very easy string handling.

8. The Whole System

This chapter should give an insight in the manner the ConStruct system works. First we introduce the configuration file with which the whole system is configured. Then we present a sample run of the system. Finally we take a view of Xbbs, the graphical front end of the ConStruct system.

The Configuration

The blackboard system can be created by giving a declarative specification of its components. This specification has to be formatted as an SGML document using the CONFIG Document Type Definition (DTD):

```
<!DOCTYPE CONFIG [
<!ELEMENT CONFIG     - - (BLACKBOARD & CONTROL & EXPERTS)>
<!ELEMENT BLACKBOARD - - (NAME & CLASS)>
<!ELEMENT CONTROL    - - (NAME & CLASS)>
<!ELEMENT EXPERTS    - - (EXPERT+)>
<!ELEMENT EXPERT     - - (NAME & CLASS & (ATTRIBUTES?))>
<!ELEMENT ATTRIBUTES - - (ATTRIBUTE+)>
<!ELEMENT ATTRIBUTE  - - (NAME & VALUE)>
<!ELEMENT NAME       - - (#PCDATA)>
<!ELEMENT CLASS      - - (#PCDATA)>
<!ELEMENT VALUE      - - (LIST | (#PCDATA))>
<!ELEMENT LIST       - - (ITEM+)>
<!ELEMENT ITEM       - - (#PCDATA)>
]>
```

According to this CONFIG DTD a valid configuration file could look like the following example (an entire configuration file is given in the appendix):

```
<CONFIG>
 <BLACKBOARD>
   <CLASS>Blackboard</CLASS>
   <NAME>Board</NAME>
 </BLACKBOARD>

 <CONTROL>
   <CLASS>Control</CLASS>
   <NAME>Basic Control System</NAME>
   <ATTRIBUTES>
     <ATTRIBUTE>
       <NAME>threshold</NAME>
       <VALUE>0.02</VALUE>
     </ATTRIBUTE>
   </ATTRIBUTES>
 </CONTROL>

 <EXPERTS>
   <EXPERT>
     <CLASS>MailExpert</CLASS>
     <NAME>Expert for RFC 822 eMails</NAME>
   </EXPERT>

   <EXPERT>
     <CLASS>NewsExpert</CLASS>
     <NAME>Expert for RFC 1036 USENET Messages</NAME>
   </EXPERT>
```

```
    <EXPERT>
      <CLASS>SignatureExpert</CLASS>
      <NAME>Expert for eMail Signatures</NAME>
      <ATTRIBUTES>
        <ATTRIBUTE>
          <NAME>maximumLines</NAME>
          <VALUE>11</VALUE>
        </ATTRIBUTE>
      </ATTRIBUTES>
    </EXPERT>

    <EXPERT>
      <CLASS>QuoteExpert</CLASS>
      <NAME>Expert for quotes</NAME>
      <ATTRIBUTES>
        <ATTRIBUTE>
          <NAME>prefixes</NAME>
          <VALUE>
            <LIST>
              <ITEM>&gt;</ITEM>
              <ITEM>:</ITEM>
            </LIST>
          </VALUE>
        </ATTRIBUTE>
      </ATTRIBUTES>
    </EXPERT>
  </EXPERTS>
</CONFIG>
```

There are three sections in the configuration file, which correspond to the three components of blackboard systems: The blackboard, the control and the experts. For each component two attributes have to be specified: The class attribute specifies the name of the class to be loaded and the name attribute gives a unique name to the component. The control system and the experts may have further attributes, where each attribute must have a name and a value.

The configuration file makes it possible to configure the whole system without having to recompile the sources.

Even if the configuration file is formatted properly, there can be some errors: The specified classes have to exist, the names have to be unique, and the specified attributes have to be known by the experts or the control system.

A sample Run

The following chapter will illustrate the working method of the whole system by describing the single steps that are done to tag an exemplary eMail.

The eMail to be examined is given below. It is an hypothetic eMail answer to a letter, which was recorded on two gold–coated phonographic records. These are carried through space by the two spacecrafts Voyager I and II.

```
From QulongFU <EOCUAP@HQ> Mon Jan 13 19:12:11 1997
Status: O
X-Status:
Date: Mon, 13 Jan 1997 19:12:07  -0500
From: QulongFU <EOCUAP@HQ>
Subject: Your letter from Mon, 13 Jan 1997a gold-coated phonographic record
Sender: QulongFU <EOCUAP@HQ>
To: Kurt Waldheim <waldheim@UN.org>
Message-ID: <Pine.HPP.3.95.970113191135.4865A-100000@pwt.ccg>
MIME-Version: 1.0
Content-Type: TEXT/PLAIN; charset=US-ASCII

On Thu, 25 Aug 1977, Kurt Waldheim wrote:

> THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
> states who represent almost all of the human inhabitants of the planet
> Earth, I send greetings on behalf of the people of our planet. We step out
> of our Solar System into the universe seeking only peace and friendship,
> to teach if we are called upon, to be taught if we are fortunate. We know
> full well that our planet and all its inhabitants are but a small part of
> the immense universe that surrounds us and it is with humility and hope
> that we take this step.
>

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.

We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.


Regards,
   QulongFU

--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
```

After putting this eMail on the blackboard, escaping the relevant characters, the initial board content looks like this:

```
<UFO>
From QulongFU &lt;EOCUAP@HQ&gt; Mon Jan 13 19:12:11 1997
Status: O
X-Status:
Date: Mon, 13 Jan 1997 19:12:07  -0500
From: QulongFU &lt;EOCUAP@HQ&gt;
Subject: Your letter from Mon, 13 Jan 1997
Sender: QulongFU &lt;EOCUAP@HQ&gt;
To: Kurt Waldheim &lt;waldheim@UN.org&gt;
Message-ID: &lt;Pine.HPP.3.95.970113191135.4865A-100000@pwt.ccg&gt;
MIME-Version: 1.0
Content-Type: TEXT/PLAIN; charset=US-ASCII

On Thu, 25 Aug 1977, Kurt Waldheim wrote:

&gt; THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
&gt; states who represent almost all of the human inhabitants of the planet
&gt; Earth, I send greetings on behalf of the people of our planet. We step out
&gt; of our Solar System into the universe seeking only peace and friendship,
&gt; to teach if we are called upon, to be taught if we are fortunate. We know
&gt; full well that our planet and all its inhabitants are but a small part of
&gt; the immense universe that surrounds us and it is with humility and hope
&gt; that we take this step.
&gt;

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.

We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.


Regards,
   QulongFU

--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
</UFO>
```

Now the control asks all the experts for their possible contributions in the order given in the configuration file. Within 13 steps, the board will be solved.

```
Step 1: Asking all Experts...

RFC 822 eMails             (1)
RFC 1036 USENET Messages   (0.5)
subdocuments
eMail Signatures           (0.1)
Paragraphs
preformatted Testareas
quotes                     (1)
Greetings                  (0.5)
Signatures                 (0.85)
emphasized Texts
pure text sections

Now engage the Expert for RFC 822 eMails to make its contribution.
```

The value in brackets indicates the returned rating for the recognized region. There can be more than one rating if an expert has more contributions to offer. The higher the rating (a value between 0 and 1), the more certain the expert is about the contribution it offers. If there is no value returned the given expert has nothing to contribute.

If no number is given, the expert has nothing to contribute. Note that every expert is asked in each step.

If two ore more ratings are equal, the first expert in the list is chosen. In the given situation the media expert for RFC 822 conform eMails is engaged by the control system.

Note that the expert for eMail Signatures now returns a value of 0.1. As soon as the eMail expert made its contribution this rating will raise.

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR>
     <NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:

&gt; THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
&gt; states who represent almost all of the human inhabitants of the planet
&gt; Earth, I send greetings on behalf of the people of our planet. We step out
&gt; of our Solar System into the universe seeking only peace and friendship,
&gt; to teach if we are called upon, to be taught if we are fortunate. We know
&gt; full well that our planet and all its inhabitants are but a small part of
&gt; the immense universe that surrounds us and it is with humility and hope
&gt; that we take this step.
&gt;

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.

We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.


Regards,
   QulongFU

--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
   </UFO>
 </BODY>
</MAIL>
```

```
Step 2: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures           (1)
Paragraphs                 (0.15)(0.7)(0.15)(0.17)(1)(0.17)
preformatted Testareas     (0.005)(0.005)(0.005)(0.005)(0.005)
                          (0.005)(0.085)
quotes                     (1)
Greetings                  (0.5)
Signatures                 (0.85)
emphasized Texts
pure text sections

Now engage the Expert for eMail Signatures to make its contribution.
```

Here is an example where an expert has recognized more than one text area. Note that the ratings differ. For instance the expert for paragraphs rates text with many lines higher than those with less. But you can see that it is irritated by the regular introduction (mail quotes) in front of each line of the largest paragraph. We will see that as soon as the quote is resolved, the rating of the paragraph expert for the interior part will increase.

Note that every paragraph has a slight chance to be recognized as a preformatted text area. But the rating is too small, so the expert will never be chosen by the control to perform its contribution as long as other experts result in a higher rating.

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR><NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:

&gt; THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
&gt; states who represent almost all of the human inhabitants of the planet
&gt; Earth, I send greetings on behalf of the people of our planet. We step out
&gt; of our Solar System into the universe seeking only peace and friendship,
&gt; to teach if we are called upon, to be taught if we are fortunate. We know
&gt; full well that our planet and all its inhabitants are but a small part of
&gt; the immense universe that surrounds us and it is with humility and hope
&gt; that we take this step.
&gt;

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.


Regards,
   QulongFU


   </UFO>
   <SIGNATURE>
--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
   </SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 3: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                   (0.15)(0.7)(0.15)(0.17)(1)(0.17)
preformatted Testareas       (0.005)(0.005)(0.005)(0.005)
                            (0.005)(0.005)
quotes                       (1)
Greetings                    (0.5)
Signatures                   (0.85)
emphasized Texts
pure text sections

Now engage the Expert for Paragraphs to make its contribution.
```

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR>
     <NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:

&gt; THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
&gt; states who represent almost all of the human inhabitants of the planet
&gt; Earth, I send greetings on behalf of the people of our planet. We step out
&gt; of our Solar System into the universe seeking only peace and friendship,
&gt; to teach if we are called upon, to be taught if we are fortunate. We know
&gt; full well that our planet and all its inhabitants are but a small part of
&gt; the immense universe that surrounds us and it is with humility and hope
&gt; that we take this step.
&gt;

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.
     </UFO>
   </PARAGRAPH>
   <UFO>

Regards,
   QulongFU


   </UFO>
   <SIGNATURE>
--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
   </SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 4: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                  (0.15)(0.7)(0.15)(0.17)(0.17)
preformatted Testareas      (0.005)(0.005)(0.005)(0.005)(0.005)
quotes                      (1)
Greetings                   (0.5)
Signatures                  (0.85)
emphasized Texts            (1)
pure text sections          (0.05)

Now engage the Expert for quotes to make its contribution.
```

Since the expert for emphasized texts works only on UFOs within paragraphs, this expert was prevented from firing until now.

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR>
     <NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:
   </UFO>
   <QUOTE>
     <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
     </UFO>
   </QUOTE>
   <UFO>

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.
     </UFO>
   </PARAGRAPH>
   <UFO>

Regards,
   QulongFU


   </UFO>
   <SIGNATURE>
       ...
   </SIGNATURE>

 </BODY>
</MAIL>
```

```
Step 5: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
               nextStep()
               RFC 822 eMails
               RFC 1036 USENET Messages
               No expert has anything to contribute... cleaning the board.
eMail Signatures
Paragraphs                   (0.15)(1)(0.15)(0.17)(0.17)
preformatted Testareas       (0.005)(0.005)(0.005)(0.005)(0.005)
quotes
Greetings                    (0.5)
Signatures                   (0.85)
emphasized Texts             (1)
pure text sections           (0.05)

Now engage the Expert for Paragraphs to make its contribution.
```

As we stated above: The quote is resolved, the rating for the interior to be a paragraph, increases from 0.7 to 1.

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR>
     <NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:

   </UFO>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>

THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system *immediately*.
     </UFO>
   </PARAGRAPH>
   <UFO>
Regards,
   QulongFU
   </UFO>
   <SIGNATURE>
       ...
   </SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 6: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                 (0.15)(0.15)(0.17)(0.17)
preformatted Testareas     (0.005)(0.005)(0.005)(0.005)
quotes
Greetings                  (0.5)
Signatures                 (0.85)
emphasized Texts           (1)
pure text sections         (0.05)(0.05)

Now engage the Expert for emphasized Texts to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:


   </UFO>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>

Regards,
   QulongFU


   </UFO>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 7: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                  (0.15)(0.15)(0.17)(0.17)
preformatted Testareas      (0.005)(0.005)(0.005)(0.005)
quotes
Greetings                   (0.5)
Signatures                  (0.85)
emphasized Texts
pure text sections          (0.05)(0.05)

Now engage the Expert for Signatures to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:


   </UFO>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>

Dear Kurt Waldheim,

we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>


   </UFO>
   <SALUT kind="Signature">
     Regards, QulongFU
   </SALUT>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 8: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                 (0.15)(0.15)(0.17)
preformatted Testareas     (0.005)(0.005)(0.005)
quotes
Greetings                  (0.5)
Signatures
emphasized Texts
pure text sections         (0.05)(0.05)

Now engage the Expert for Greetings to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:


   </UFO>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>


   </UFO>
   <SALUT kind="Greeting">
     Dear Kurt Waldheim,
   </SALUT>
   <UFO>
we intercepted your unmanned vehicle interfering with our sovereign
territory.


   </UFO>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>


   </UFO>
   <SALUT kind="Signature">
     Regards, QulongFU
   </SALUT>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 9: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                  (0.15)(0.17)
preformatted Testareas      (0.005)(0.005)
quotes
Greetings
Signatures
emphasized Texts
pure text sections          (0.05)(0.05)

Now engage the Expert for Paragraphs to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:


   </UFO>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>


   </UFO>
   <SALUT kind="Greeting">
     Dear Kurt Waldheim,
   </SALUT>
   <PARAGRAPH justify="blockoleft">
     <UFO>
we intercepted your unmanned vehicle interfering with our sovereign
territory.
     </UFO>
   </PARAGRAPH>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>


   </UFO>
   <SALUT kind="Signature">
     Regards, QulongFU
   </SALUT>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

The justify attribute of the paragraph is called blockoleft. This means that the expert is not able to determine if the paragraph is justified left or justified block. This is because the content of the paragraph is only one line (in the original document).

```
Step 10: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs                 (0.15)
preformatted Testareas     (0.005)
quotes
Greetings
Signatures
emphasized Texts
pure text sections         (0.05)(0.05)(0.05)

Now engage the Expert for Paragraphs to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <PARAGRAPH justify="blockoleft">
     <UFO>
On Thu, 25 Aug 1977, Kurt Waldheim wrote:
     </UFO>
   </PARAGRAPH>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>


   </UFO>
   <SALUT kind="Greeting">Dear Kurt Waldheim,</SALUT>
   <PARAGRAPH justify="blockoleft">
     <UFO>
we intercepted your unmanned vehicle interfering with our sovereign
territory.
     </UFO>
   </PARAGRAPH>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>


   </UFO>
   <SALUT kind="Signature">Regards, QulongFU</SALUT>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

```
Step 11: Asking all Experts...

RFC 822 eMails
RFC 1036 USENET Messages
subdocuments
eMail Signatures
Paragraphs
preformatted Testareas
quotes
Greetings
Signatures
emphasized Texts
pure text sections          (0.05)(0.05)(0.05)(0.05)

Now engage the Expert for pure text sections to make its contribution.
```

```
<MAIL>
 <HEAD>...</HEAD>
 <BODY>
   <PARAGRAPH justify="blockoleft">
     On Thu, 25 Aug 1977, Kurt Waldheim wrote:
   </PARAGRAPH>
   <QUOTE>
     <PARAGRAPH justify="left">
       <UFO>
THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147 member
states who represent almost all of the human inhabitants of the planet
Earth, I send greetings on behalf of the people of our planet. We step out
of our Solar System into the universe seeking only peace and friendship,
to teach if we are called upon, to be taught if we are fortunate. We know
full well that our planet and all its inhabitants are but a small part of
the immense universe that surrounds us and it is with humility and hope
that we take this step.
       </UFO>
     </PARAGRAPH>
   </QUOTE>
   <UFO>


   </UFO>
   <SALUT kind="Greeting">Dear Kurt Waldheim,</SALUT>
   <PARAGRAPH justify="blockoleft">
     <UFO>
we intercepted your unmanned vehicle interfering with our sovereign
territory.
     </UFO>
   </PARAGRAPH>
   <PARAGRAPH justify="left">
     <UFO>
We must insist on the retraction of this space-junk.  According to
our law you are obliged to take care of your property and are
responsible for every harm arises thereof! In the case of a repeat
offense we will be forced to shutdown your solar system
     </UFO>
     <EMPHASIZED style="bold">immediately</EMPHASIZED>
     <UFO>
.
     </UFO>
   </PARAGRAPH>
   <UFO>


   </UFO>
   <SALUT kind="Signature">Regards, QulongFU</SALUT>
   <SIGNATURE>...</SIGNATURE>
 </BODY>
</MAIL>
```

...the Expert for pure text sections makes its contributions several times.

The final board looks like:

```
<MAIL>
 <HEAD>
   <SUBJECT>Your letter from Mon, 13 Jan 1997</SUBJECT>
   <AUTHOR>
     <NAME>QulongFU</NAME>
     <EMAIL-ADDRESS>EOCUAP@HQ</EMAIL-ADDRESS>
   </AUTHOR>
   <ADDRESSEE>
     <NAME>Kurt Waldheim</NAME>
     <EMAIL-ADDRESS>waldheim@UN.org</EMAIL-ADDRESS>
   </ADDRESSEE>
 </HEAD>
 <BODY>
   <PARAGRAPH justify="blockoleft">
     On Thu, 25 Aug 1977, Kurt Waldheim wrote:
   </PARAGRAPH>
   <QUOTE>
     <PARAGRAPH justify="left">
       THE SECRETARY GENERAL OF THE UNITED Nations, an organization of 147
       member states who represent almost all of the human inhabitants of the
       planet Earth, I send greetings on behalf of the people of our planet.
       We step out of our Solar System into the universe seeking only peace
       and friendship, to teach if we are called upon, to be taught if we are
       fortunate. We know full well that our planet and all its inhabitants
       are but a small part of the immense universe that surrounds us and it
       is with humility and hope that we take this step.
     </PARAGRAPH>
   </QUOTE>
   <SALUT kind="Greeting">
     Dear Kurt Waldheim,
   </SALUT>
   <PARAGRAPH justify="blockoleft">
     we intercepted your unmanned vehicle interfering with our sovereign
     territory.
   </PARAGRAPH>
   <PARAGRAPH justify="left">
     We must insist on the retraction of this space-junk. According to our
     law you are obliged to take care of your property and are responsible
     for every harm arises thereof! In the case of a repeat offense we will
     be forced to shutdown your solar system <EMPHASIZED style="bold">
     immediately</EMPHASIZED>.
   </PARAGRAPH>
   <SALUT kind="Signature">
     Regards, QulongFU
   </SALUT>
   <SIGNATURE>
--
      QulongFU (EOCUAP assistant)
Email: EOCUAP@HQ
UWW:   http://uww.EOCUAP.org/~QulongFU/index.sgml
   </SIGNATURE>
 </BODY>
</MAIL>
```

Xbbs

We implemented a graphical front end called Xbbs for our blackboard system. With this tool the configuration of the system should have been much easier than doing so with command line options. We did not manage to implement all the features we wanted to, so this tool is only capable to read and write files and solve the board at once or in single step mode.

9. Résumé

Java

Java – as an object–oriented programming language – was adequate for implementing the blackboard system as a whole and the SGML class. It was not suitable for writing the experts regarding the problems of structure recognition. This is mainly because Java lacks of libraries that would relieve from the task of finding certain text elements and their corresponding coordinates.

We used a library offering regular expressions and the vast amount of methods to manipulate strings with Java. Nonetheless the implementation of experts turned out to be quite cumbersome. The reason for this is mainly that we had no construct to handle coordinates and use regular expressions in an easy way. The design and implementation of a meta–language to implement the experts in a more declarative way would be a very useful extension of our work.

Java doesn't provide higher–order functions. Those would accept functions as parameters and/or return functions as return values. Higher–order functions provide a high degree of modularity and code reuse. All of our experts work on a tree–like SGML data structure. With higher–order functions the code sequences to run through these structures could be easily shared.

Java also does not provide enumeration types. They can be simulated, but this is only a way around and not a proper solution. We need enumeration types, for example, to construct our SGML data types (CDATA, RCDATA, PCDATA, ...).

The last feature we would have needed and that is not supported by Java are parameterized types and modules. These are necessary to implement lists of a certain kind without proving 'by hand' that the added element is a type or subtype of the intended class.

Last but not least Java has longer turn–around times than interpreted languages like Perl. This significantly slowed down the implementation times of the experts and was an obstacle to our goal to build the whole system by rapid prototyping.

The advantages of Java are undoubtedly its platform independence and its wide acceptance. A useful tool is javadoc, which generates an online and a printable documentation from the code. Since the tags for the javadoc program are part of Java's specification, this kind of literate programming is widely accepted by Java programmers.

Despite the unbalance of the advantages and disadvantages, we do not regret using Java.

Blackboard System

The blackboard architecture proved to be very capable of solving the task of structure recognition. The numerous experts could easily be developed by different programmers at the same time. If there are some structures not yet recognized, then a new expert can be added to adapt the system without changing the existing parts.

Because of the modularity of blackboard systems it is also effortless to exchange the control and the blackboard.

Our implementation was done to prove the work–ability of a blackboard system for the task of structure recognition in texts. We had no eye on efficiency, but we tried to make it easy to extend the existing code. Therefore a lot of improvement is possible; especially the control, but also the blackboard and the experts could get some enhancements. For a practical use of the ConStruct system, some additional experts are missing (expert for lists, headlines, source code sequences, ...) and some existing ones should be improved to recognize wider variation of formattings.

10. Appendix

An entire configuration file

```
<CONFIG>
 <BLACKBOARD>
   <CLASS>Blackboard</CLASS>
   <NAME>Board</NAME>
 </BLACKBOARD>

 <CONTROL>
   <CLASS>Control</CLASS>
   <NAME>Basic Control System</NAME>
   <ATTRIBUTES>
     <ATTRIBUTE>
       <NAME>threshold</NAME>
       <VALUE>0.02</VALUE>
     </ATTRIBUTE>
   </ATTRIBUTES>
 </CONTROL>

 <EXPERTS>
   <EXPERT>
     <CLASS>MailExpert</CLASS>
     <NAME>Expert for RFC 822 eMails</NAME>
   </EXPERT>

   <EXPERT>
     <CLASS>NewsExpert</CLASS>
     <NAME>Expert for RFC 1036 USENET Messages</NAME>
   </EXPERT>

   <EXPERT>
     <CLASS>SubDocExpert</CLASS>
     <NAME>Expert for subdocuments</NAME>
     <ATTRIBUTES>
       <ATTRIBUTE>
         <NAME>configfile</NAME>
         <VALUE>../cfg/SubDoc.config</VALUE>
       </ATTRIBUTE>
       <ATTRIBUTE>
         <NAME>repository</NAME>
         <VALUE>.</VALUE>
       </ATTRIBUTE>
     </ATTRIBUTES>
   </EXPERT>

   <EXPERT>
     <CLASS>SignatureExpert</CLASS>
     <NAME>Expert for eMail Signatures</NAME>
     <ATTRIBUTES>
       <ATTRIBUTE>
         <NAME>maximumLines</NAME>
         <VALUE>11</VALUE>
       </ATTRIBUTE>
     </ATTRIBUTES>
   </EXPERT>

   <EXPERT>
     <CLASS>ParagraphExpert</CLASS>
     <NAME>Expert for Paragraphs</NAME>
   </EXPERT>

   <EXPERT>
     <CLASS>PreformatExpert</CLASS>
     <NAME>Expert for preformatted Testareas</NAME>
     <ATTRIBUTES>
       <ATTRIBUTE>
         <NAME>factor</NAME>
```

```
        <VALUE>0.5</VALUE>
      </ATTRIBUTE>
    </ATTRIBUTES>
  </EXPERT>

  <EXPERT>
    <CLASS>QuoteExpert</CLASS>
    <NAME>Expert for quotes</NAME>
    <ATTRIBUTES>
      <ATTRIBUTE>
        <NAME>prefixes</NAME>
        <VALUE>
          <LIST>
            <ITEM>&gt;</ITEM>
            <ITEM>:</ITEM>
          </LIST>
        </VALUE>
      </ATTRIBUTE>
    </ATTRIBUTES>
  </EXPERT>

  <EXPERT>
    <CLASS>SalutExpert</CLASS>
    <NAME>Expert for Greetings</NAME>
    <ATTRIBUTES>
      <ATTRIBUTE>
        <NAME>kind</NAME>
 <VALUE>Greeting</VALUE>
      </ATTRIBUTE>
      <ATTRIBUTE>
        <NAME>UnrecognizedWordsToQuit</NAME>
 <VALUE>4</VALUE>
      </ATTRIBUTE>
      <ATTRIBUTE>
        <NAME>AverageSalutationLineLength</NAME>
 <VALUE>60</VALUE>
      </ATTRIBUTE>
      <ATTRIBUTE>
        <NAME>salutations</NAME>
        <VALUE>
          <LIST>
            <ITEM>Lieber</ITEM>
            <ITEM>Liebe</ITEM>
            <ITEM>Liebes</ITEM>
            <ITEM>Hallo</ITEM>
            <ITEM>Hi</ITEM>
            <ITEM>Hello</ITEM>
            <ITEM>Greetings</ITEM>
            <ITEM>Dear</ITEM>
          </LIST>
        </VALUE>
      </ATTRIBUTE>
      <ATTRIBUTE>
        <NAME>names</NAME>
        <VALUE>
          <LIST>
            <ITEM>François</ITEM>
            <ITEM>Bry</ITEM>
            <ITEM>Norbert</ITEM>
            <ITEM>Eisinger</ITEM>
            <ITEM>Tim</ITEM>
            <ITEM>Geisler</ITEM>
            <ITEM>Thomas</ITEM>
            <ITEM>Traber</ITEM>
            <ITEM>Robert</ITEM>
            <ITEM>Reiner</ITEM>
          </LIST>
        </VALUE>
      </ATTRIBUTE>
    </ATTRIBUTES>
  </EXPERT>
```

```
   <EXPERT>
     <CLASS>SalutExpert</CLASS>
     <NAME>Expert for Signatures</NAME>
     <ATTRIBUTES>
       <ATTRIBUTE>
  <NAME>kind</NAME>
  <VALUE>Signature</VALUE>
        </ATTRIBUTE>
       <ATTRIBUTE>
         <NAME>UnrecognizedWordsToQuit</NAME>
  <VALUE>4</VALUE>
       </ATTRIBUTE>
       <ATTRIBUTE>
         <NAME>AverageSalutationLineLength</NAME>
  <VALUE>60</VALUE>
       </ATTRIBUTE>
        <ATTRIBUTE>
         <NAME>salutations</NAME>
         <VALUE>
           <LIST>
             <ITEM>Gruss</ITEM>
             <ITEM>Gruß</ITEM>
             <ITEM>Greetings</ITEM>
             <ITEM>Regards</ITEM>
             <ITEM>Best regards</ITEM>
             <ITEM>Bye</ITEM>
           </LIST>
         </VALUE>
       </ATTRIBUTE>
       <ATTRIBUTE>
         <NAME>names</NAME>
         <VALUE>
           <LIST>
             <ITEM>François</ITEM>
             <ITEM>Bry</ITEM>
             <ITEM>Norbert</ITEM>
             <ITEM>Eisinger</ITEM>
             <ITEM>Tim</ITEM>
             <ITEM>Geisler</ITEM>
             <ITEM>Thomas</ITEM>
             <ITEM>Traber</ITEM>
             <ITEM>Robert</ITEM>
             <ITEM>Reiner</ITEM>
           </LIST>
         </VALUE>
       </ATTRIBUTE>
     </ATTRIBUTES>
   </EXPERT>

   <EXPERT>
     <CLASS>EmphasizeExpert</CLASS>
     <NAME>Expert for emphasized Texts</NAME>
   </EXPERT>

   <EXPERT>
     <CLASS>TextExpert</CLASS>
     <NAME>Expert for pure text sections</NAME>
   </EXPERT>
 </EXPERTS>
</CONFIG>
```

11. Bibliography

[1]

Ken Arnold and James Gosling:
The Java Programming Language,
Addison–Wesley 1996,
ISBN 0–201–63455–4.

[2]

Grady Booch,
Object Oriented Design: With Applications
Benjamin/Cummings, 1991
ISBN 0–8053–0091–0
(see especially chapter 11: *Common Lisp Object System: Cryptanalysis*)

[3]

Charles Goldfarb
SGML Handbook
Oxford University Press, 1990
ISBN 0–19–853737–9

[4]

International Standard Organization
ISO 8879: Information processing – Text and office systems – Standard Generalized Markup Language (SGML)

[5]

Martin Bryan
SGML – An Author's Guide to the Standard Generalized Markup Language
Addison–Wesley, 1988
ISBN 0–201–17535–5

[6]

Martin Colby and David S. Jackson
Using SGML
Que, 1996
ISBN 0–7897–0414–5

[7]

Lee Erman, Frederick Hayes–Roth, Victor Lesser, and D. Raj Reddy,
The Hearsay–II Speech–Understanding System: Integrating Knowledge to Resolve Uncertainty
Computer Surveys, vol. 12, no. 2, 213–253, 1980
(also in *Blackboard Systems*, Robert Engelmore and Tony Morgan, editors, Addison–Wesley, 31–86, 1988).

[8]

Norman Carver and Victor Lesser
The Evolution of Blackboard Control Architectures
CMPSCI Technical Report 92–71
October 1992
Page 3

[9]

Englemore, R. and Morgan, T.
Blackboard Systems
Addison–Wesley, 1988,
Page 6

[10]

Nii, P.
Blackboard Systems: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures.
AI Magazine, vol. 7 (2), Summer 1986,
Page 43.

[11]

David H. Crocker
RFC 822: Standard For The Format Of ARPA Internet Text Messages