Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

Research
Report
RR-91-06

# Synthesizing Illustrated Documents:

# A Plan-Based Approach

**Elisabeth André, Thomas Rist**

**January 1991**

**Deutsches Forschungszentrum für Künstliche Intelligenz GmbH**

Postfach 20 80
D-6750 Kaiserslautern, FRG
Tel.: (+49 631) 205-3211/13
Fax: (+49 631) 205-3210

Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, FRG
Tel.: (+49 681) 302-5252
Fax: (+49 681) 302-5341

# Deutsches Forschungszentrum
# für
# Künstliche Intelligenz

The German Research Center for Artificial Intelligence (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI) with sites in Kaiserslautern und Saarbrücken is a non-profit organization which was founded in 1988 by the shareholder companies ADV/Orga, AEG, IBM, Insiders, Fraunhofer Gesellschaft, GMD, Krupp-Atlas, Mannesmann-Kienzle, Nixdorf, Philips and Siemens. Research projects conducted at the DFKI are funded by the German Ministry for Research and Technology, by the shareholder companies, or by other industrial contracts.

The DFKI conducts application-oriented basic research in the field of artificial intelligence and other related subfields of computer science. The overall goal is to construct *systems with technical knowledge and common sense* which - by using AI methods - implement a problem solution for a selected application area. Currently, there are the following research areas at the DFKI:

- Intelligent Engineering Systems
- Intelligent User Interfaces
- Intelligent Communication Networks
- Intelligent Cooperative Systems.

The DFKI strives at making its research results available to the scientific community. There exist many contacts to domestic and foreign research institutions, both in academy and industry. The DFKI hosts technology transfer workshops for shareholders and other interested groups in order to inform about the current state of research.

From its beginning, the DFKI has provided an attractive working environment for AI researchers from Germany and from all over the world. The goal is to have a staff of about 100 researchers at the end of the building-up phase.

Prof. Dr. Gerhard Barth
Director

# Synthesizing Illustrated Documents: A Plan-Based Approach

**Elisabeth André, Thomas Rist**

DFKI-RR-91-06

This paper has also been published in the Proceedings of InfoJapan`90.

# Synthesizing Illustrated Documents: A Plan-Based Approach

Elisabeth André, Thomas Rist

German Research Center for Artificial Intelligence
Saarbrücken Site, Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, Germany

**Abstract:**

The aim of our work* is to develop a system able to generate documents in which text and pictures are smoothly integrated. Such tailoring requires knowledge concerning the functions of textual and pictorial document parts and the relations between them. We start from the assumption that not only the generation of text, but also the generation of multimodal documents can be considered as a sequence of communicative acts which aim to achieve certain goals. Based on textlinguistic work, the structure of an illustrated document is described by the hierarchical order of communicative acts and the relations between them. In view of the generation of text-picture combinations, we have examined relations which frequently occur between text passages and pictures, or between the parts of a picture. For the automated generation of illustrated documents, we propose a plan-based approach. To represent knowledge about presentation techniques, we have designed presentation strategies which relate to both text and picture production. Finally, we show by example how a document fragment is synthesized.



---

* The research presented here has been carried out in the WIP project which is supported by the German Ministry for Research and Technology under ITW8901 8. Special thanks to Wolfgang Wahlster for his valuable remarks on earlier versions of this paper.

# 1 Introduction

The effective application of information systems, help systems, control panels, or expert systems of the next generation requires intelligent user interfaces able to present information in a flexible manner appropriate to various presentation situations (cf. [Wahlster et al. 89]). There seems to be general agreement that in these interfaces the presentation of information should incorporate different modes, particularly text and graphics. The most important motivation for using text-picture combinations is the fact that in many situations information can be conveyed more precisely and efficiently through illustrated documents than through the exclusive use of either text or graphics. However, the use of text-embedded pictures is not advantageous per se, eg., see the experimental findings of text-picture researchers discussed in [Levie 87] and [Molitor et al. 89], or think of your own experience with badly crafted manuals for maintenance and repair in which illustrations, rather than facilitating comprehension, may even confuse the user.

The aim of our work is to develop a system able to generate documents in which text and pictures are smoothly integrated. It is one thing to simply merge verbalization and visualization results into a multimodal output, but quite another to generate presentations where textual and pictural parts are tailored to each other. Therefore, such a system has to decide carefully when to use graphics/text in place of text/graphics and when to use a mixture of both modes.

# 2 The Structure of Illustrated Documents

Our approach is based on the assumption that not only the generation of text, but also the generation of multimodal documents can be considered as an act sequence which aims to achieve certain goals. We presume that there is at least one act which is central to the goal of the whole document. This act is referred to as the *main act*. Acts supporting the main act are called *subsidiary acts*.[1] Since main and subsidiary acts can, in turn, be composed of main and subsidiary acts, a hierarchical document structure results. While the root of the hierarchy generally corresponds to a complex communicative act such as describing a process, the leaves are elementary acts, i.e., speech acts (cf. [Searle 69]) or pictorial acts (cf. [Kjorup 78]).

The structure of a document is, however, not only determined by its hierarchical act structure, but also by the role acts play in relation to other acts. In textlinguistic studies, a variety of coherence relations between text segments has been proposed (e.g., see [Grimes 75] and [Hobbs 78]). Perhaps the most elaborated set is presented in RST-theory (cf. [Mann&Thompson 87]). Examples of RST-relations are *Motivation*, *Elaboration*, *Enablement*, *Interpretation* or *Summary*. Text-picture researchers have investigated the role a particular picture plays in relation to the accompanying text passage. E.g., Levin has found five primary functions (cf. [Levin et al. 87]): *Decoration*, *Representation*, *Organization*, *Interpretation* and *Transformation*. Hunter and colleagues distinguish between: *Embellish*, *Reinforce*,

---

[1] This distinction between main and subsidiary acts essentially corresponds to the distinction between *global* and *subsidiary speech acts* in [Searle 69], *main speech acts* and *subordinate speech acts* in [Van Dijk 80], *dominierenden Handlungen* and *subsidiären Handlungen* in [Brandt et el. 83] and between *nucleus* and *satellites* in the RST-Theory proposed in [Mann & Thompson 87].

*Elaborate*, *Summarize* and *Compare* (cf. [Hunter et al. 87]). An attempt at a transfer of the relations proposed by Hobbs to pictures and text-picture combinations has been made in [Bandyopadhyay 90]. Unfortunately, text-picture researchers only consider the communicative functions of whole pictures, i.e., they do not address the question of how a picture is organized. To get an informative description of the whole document structure, one has to consider relations between picture parts or between picture parts and text passages too. E.g., a portion of a picture can serve as background for the rest of the picture or a text passage can elaborate on a particular section of a picture.

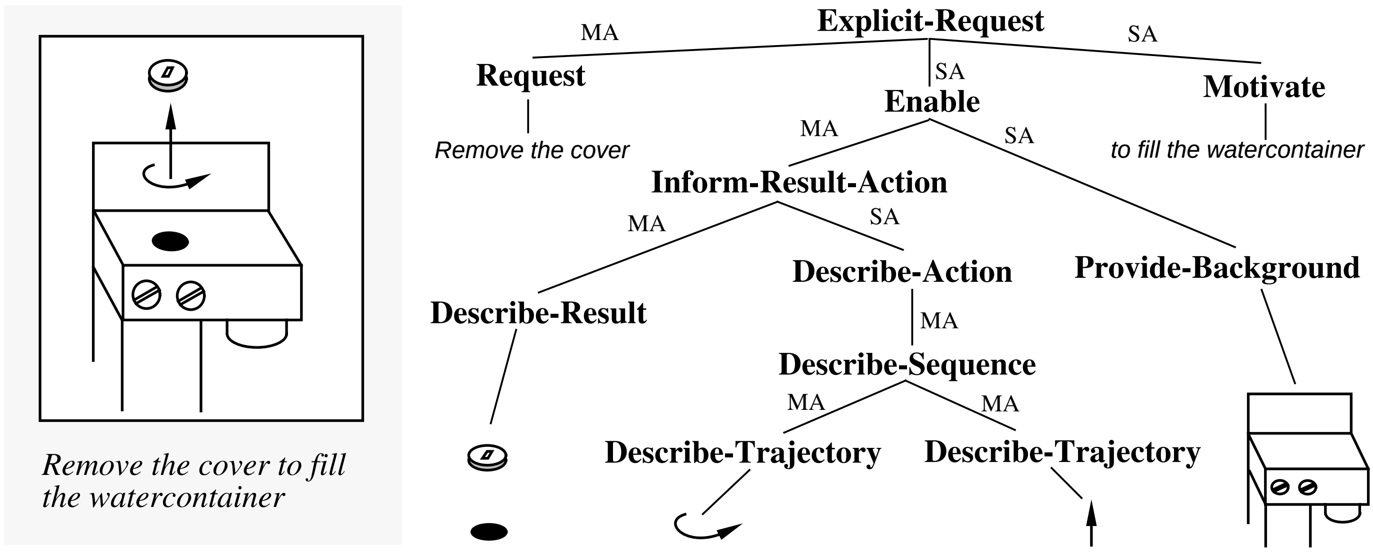


Fig. 1: A Document Fragment[2] and its Structure

To illustrate how a document containing pictures is organized, we analyze the document fragment in Fig. 1. The goal of this document fragment is to explicitly request the user to remove the cover of the watercontainer of an espresso-machine. The actual request is conveyed through text (main act). To motivate that request, the goal, i.e., the filling of the watercontainer, is mentioned (subsidiary act). The picture provides additional information which enables the addressee to carry out the request (subsidiary act). The generation of the picture is also subdivided into a main act, which describes the result and the actions to be performed, and a subsidiary act, which provides the background to facilitate orientation. The right side of Fig. 1 summarizes the hierarchical act structure of the document fragment.

## 3 Design Criteria for Text-Picture Combinations

When designing an illustrated document, an author has to decide which mode combination is the most suitable for meeting his goals. The decision-making process for mode selection is influenced by different factors. These include: the kind of information content, the communicative functions that textual and pictorial document parts ought to fill in a presentation, resource limitations (e.g., limitations due to the

---

[2] The examples in this paper are slightly modified and translated versions of the instructions for the Philips espresso-machine HD 5649.

output medium, or space and time restrictions) and user characteristics (e.g., trained or untrained in reading visual displays). In this paper, the first and the second factors are discussed in more detail.

## 3.1 Mode Preferences for Information Types

Given a certain information content, we first have to check which mode of presentation can be used. E.g., it is very difficult or even impossible to graphically depict quantifiers (such as *some* or *a few*) and modalities (such as *want*, *belief* or *necessity*). In cases where text as well as graphics may be employed, the question of which mode conveys the information most effectively arises. Several classifications of information content which seem relevant for selecting the mode of presentation have been proposed (cf. [Bieger&Glock 84], [Roth&Mattis 90] and [Feiner&McKeown 90]). In the following, we present some classification criteria which are of importance in the domain of maintenance and repair instructions for technical devices; further criteria may be necessary, in particular when shifting to another domain.

**Concrete information:** Information concerning visual properties of concepts (such as shape, color and texture) is classified as concrete. We regard events and actions as concrete if they involve physical objects and if their occurrence causes visually perceptible changes. Since pictures seem to be superior in teaching perceptual concepts (e.g., see [Molitor et al. 89]), graphics will be used in preference to text when presenting concrete information.

**Spatial information:** Since space is conceptualized mainly through objects, the category of spatial information primarily includes information concerning the location, orientation and composition of objects. Furthermore, events and actions can have a spatial component. Since a movement of a physical object can be characterized by means of spatial concepts (such as the direction of movement or the starting and end position), actions and events also get the attribute *spatial* if they involve movements of physical objects. In deciding how to present spatial information, we can partly fall back on empirical psychological studies. E.g., Bieger and Glock (cf. [Bieger&Glock 86]) found out that in assembly instructions spatial information is perceived faster if pictures are used; on the other hand, subjects confronted with textual presentations make fewer mistakes when carrying out instructions. Thus, if the emphasis of presentation is on speed, pure pictorial presentations of spatial information should be preferred.

**Temporal information:** In the domain of operating instructions, the temporal relations between states, events or actions play an important part. The sequential order of events can be effectively communicated by arranging pictures from top to bottom or from left to right. In some cases, subsequent events can be depicted in a single picture (cf. Fig. 1). While precedence relations can be easily communicated through pictures, the fact that two events overlap in time is hard to express pictorially. Furthermore, for a number of time specifications, such as *mostly* or *periodically*, pure textual presentations should be preferred in order to avoid misconceptions.

**Covariant information:** Covariant information expresses a semantic relationship between at least two pieces of information that vary together. Such relationships are: *cause/effect*, *action/result*, *problem/solution*, *condition*, and *concession*.[3] Cause/effect and action/result relationships are often expressed through a single picture, a sequence of pictures or through a text-picture combination. The presenter has, however, to consider that cause/effect and action/goal relationships between (parts of) pictures are often interpreted as pure temporal relationships. If it is not certain whether the addressee recognizes the intended relationship, text should be used in preference to graphics. To ensure that a problem/solution relationship is correctly interpreted, the problem should be presented in text unless a kind of picture language is used (e.g., in [Strothotte&Schmid 90], a question mark indicates that a picture presents a problem.). Condition and concession relationships can hardly be expressed by graphics without verbal comments.

## 3.2 Achievement of Communicative Goals

As discussed in section 2, pictures can accomplish certain communicative functions with respect to text and vice versa. Among other uses, pictures and text may be used to define, describe, inform, show, elucidate, motivate, enable, elaborate, summarize, provide a background, warn, prove, compare, attract attention or to persuade. For the design of illustrated documents, we are especially interested in those communicative functions that appear frequently in text-picture combinations. Such functions can be found in text-picture research or can be derived from the pragmatic relations between pieces of discourse set up in textlinguistic studies, e.g., in RST-theory. In the following, some of these functions are presented.

**Attract-Attention:** The text directs the addressee's attention to special aspects of the picture/text. E.g. directives, such as "Look at ..." can be used to tell the addressee what is important in a picture. Furthermore, a part of a picture can emphasize other document parts, e.g., think of arrows pointing to important objects.[4]

**Compare:** Two document parts provide a comparison between several concepts. To emphasize the differences or parallels between the concepts, the same presentation modes should be used for describing the concepts.

**Elaborate:** One part of a document provides further details about another part. Text can elaborate on a picture, e.g., by specifying attributes of an object shown in the picture. On the other hand, a picture can elaborate on text, e.g., by showing an object belonging to a verbally described class. Pictures can also elaborate on other pictures, e.g., think of an inset that shows further details of a depicted object.

**Enable:** The picture/text provides additional information in order to enable the addressee to perform the requested action. E.g., a request may be accompanied by a picture showing how an action should be carried out. The request is either implicit or verbally expressed.

---

[3] These relations also appear in RST-theory to describe a semantic relation between real world entities.

[4] This situation must not be confused with situations where a document part attracts the addressee's attention because of its visual appearence (e.g., because of its size, position or color).

**Elucidate:** One document part provides an explanation or interpretation of another part. E.g., text can be used to express the meaning of a picture or to clarify graphical techniques. While text can explain pictures or text passages, pictures can explain text, but normally not other pictures (cf. [Muckenhaupt 86]).

**Motivate:** The addressee is to be motivated to comply with a request. This goal can be met by means of pictures or by means of text. Think of an advertisement showing a cup of steaming coffee to motivate people to buy this coffee. The request follows from the context or is conveyed through text.

**Prove:** The picture/text produces evidence for a verbal claim. Since pictures increase authenticity (cf. [Smith&Smith 66]), they are well suited to support a claim.

**Provide-Background:** One document part establishes the context for the other. E.g., text may provide the necessary background information for a picture that shows a device from an extraordinary perspective. Background can also be provided by parts of a picture, e.g., a picture of an object may include further objects in order to reduce ambiguities by showing the object's spatial context.

**Summarize:** The picture/text provides an organized, reduced form of the text structure. E.g., a picture may be presented in advance to show the most important parts of a machine which are described in detail by text. On the other hand, text may be used to summarize the content of a picture.

## 4 Planning Illustrated Documents with Presentation Strategies

We start from the assumption that the author of a document has a set of presentation strategies at his disposal which can be selected and combined according to the task. Such strategies reflect general presentation knowledge as indicated in section 3 or they embody more specific knowledge of how to present a certain subject. To represent the strategies, we follow the approach proposed in [Moore&Paris 89] and [Moore&Swartout 89] to operationalize RST-theory for the planning of texts.

A strategy consists of a name, a header, an effect, a set of applicability conditions, and the main and subsidiary acts. The effect of a strategy refers to an intentional goal, e.g., *the addressee knows a particular concept*[5]. To represent intentional goals, the modal operators Goal, Bel and BMB (cf. [Cohen&Levesque 85]) are used as in Hovy's RST planner (cf. [Hovy 88]). The expression (Goal P x) stands for: The presenter P has x as goal. (Bel P x) means: P believes that x is satisfied. (BMB P A x) is an abbreviation for the infinite conjunction: (Bel P x) & (Bel P (Bel A x)) & (Bel P (Bel A (Bel P x))), etc. The header indicates the communicative function of a document part (e.g., to motivate). The applicability conditions specify in which situation a certain strategy can be used and put restrictions on the variables to be instantiated. The main and subsidiary acts form the kernel of the presentation strategies. Following textlinguistics studies (cf. [Brandt et al. 83]), we also permit strategies which contain no main act and allow for the combination of several main acts in a communicative unit if none of them is subordinated. The mode of presentation is indicated in the header. It is derived from the mode of the

[5] In [Moore&Paris 89], the effect may also be a rhetorical relation.

main acts. Depending on whether the main acts are realized in text, graphics or both modes, the values TEXT, GRAPHICS or MIXED are assigned. The mode remains unspecified until mode decisions are made for the main acts of a strategy. In the following, some presentation strategies are listed:

[1] **Name:**
Explicit-Request
**Header:**
(Explicit-Request P A ?action TEXT)
**Effect:**
(BMB P A (Goal P (Done A ?action)))
**Applicability Conditions:**
(Goal P (Done A ?action))
**Main Acts:**
(Request P A ?action)
**Subsidiary Acts:**
(Motivate P A ?action ?mode-1)
(Enable P A ?action ?mode-2)

Strategy 1 can be used to explicitly ask an addressee to perform an action. The first subsidiary act serves as a motivation; the second provides information which enables him to carry out the action. Since an action can be motivated and enabled by text, graphics or a mixture of both, the mode is not specified in the strategy.

[2] **Name:**
Enable-by-Describing-Result-Action
**Header:**
(Enable P A ?action ?mode)
**Effect:**
(BMB P A (Causes ?action ?result))
**Applicability Conditions:**
(Bel P (Causes ?action ?result))
**Main Acts:**
(Inform P A (Causes ?action ?result) ?mode)[6]

To enable the execution of an action, the addressee should be informed about what he has to do and which result is desired (cf. Strategy 2). This information can be conveyed using Strategy 3 which provides for describing the result in a main act and the action in a subsidiary act. Inform and Describe refer to complex communicative acts which can be refined into elementary textual or pictorial acts, such as Assert or Depict.

[3] **Name:**
Inform-Result-Action
**Header:**
(Inform P A (Causes ?action ?result) ?mode-1)
**Effect:**
(BMB P A (Causes ?action ?result))
**Applicability Conditions:**
(Bel P (Causes ?action ?result))
**Main Acts:**
(Describe P A (Result ?result) ?mode-1)
**Subsidiary Acts:**
(Describe P A (Action ?action) ?mode-2)

---

[6] In an earlier version of our system, the mode had been specified already in this place. The disadvantage of this method is that it bears combinatorical problems. For each mode combination, we had to specify an individual strategy.

The result of an action can be described by describing all changes caused by the action. As an example, we look at a change of orientation. Such changes can be described by showing the objects involved in the action with their new orientation (cf. Strategy 4). When showing a picture of an object, we have to ensure that the addressee relates the picture objects to the same world objects as the presenter. If the picture is ambiguous, Strategy 5 can be applied to depict the object as a part of a larger object. The optional subsidiary act in Strategy 4 will only be carried out when enough space is available. As in Strategy 5, additional objects are depicted to facilitate the spatial orientation of the addressee.

```
[4]  Name:
       Show-Orientation
     Header:
       (Describe P A (Orientation ?orientation ?x)  GRAPHICS)
     Effect:
       (BMB P A (Orientation ?orientation ?x))
     Main Acts:
       (Depict P A (Orientation ?orientation ?x) ?px  ?picture)
     Subsidiary Acts:
       (Achieve P (BMB P A (Identifiable-P A ?x ?px ?picture)) ?mode-1)
       (*opt* (Bel P Space-Available-P)
              (Achieve P (Fac-Orientation A ?x ?px ?picture) ?mode-2))
```

```
[5]  Name:
       Enable-Identification-by-Background
     Header:
       (Provide-Background P A ?x ?px ?picture GRAPHICS)
     Effect:
       (BMB P A (Identifiable-P A ?x ?px ?picture))
     Applicability Conditions:
       (AND (Bel P (Perceptually-Accessible-P A ?x))
            (Bel P (Part-of ?x ?z)))
     Main Acts:
       (Depict P A (Background ?z)  ?pz ?picture)
     Subsidiary Acts:
       (Achieve P (BMB P A (Identifiable-P A ?z ?pz ?picture)) ?mode)
```

To automatically generate documents, the presentation strategies are treated as operators of a planning system. During the planning process, the planner tries to find strategies which are either specified by the header or whose effect matches the presentation goal. The planner then checks whether their applicability conditions hold. If several strategies are applicable, metarules are applied, cf. 6. Among other things, these metarules embody information concerning specific mode preferences and the extent to which different modes can be combined. The criteria discussed in section 3 can be used as arguments voting for or against a particular mode of presentation for a certain information content in a specific situation. These arguments are collected and weighed against each other.

```
[6]   IF (IS-A  ?current-attribute-value Spatial-Concept)
        THEN (Dobefore *graphics-strategies* *text-strategies*)
```

After the selection of a strategy, the corresponding main and subsidiary acts are carried out unless the corresponding presentation goals are already satisfied. Whether the main acts are performed before or after the subsidiary acts is conditional upon their dependency on the subsidiary actions. E.g., to elucidate a picture, knowledge about the picture is necessary. In this case, the picture has to be produced first. After that, it can be examined by anticipating the addressee's understanding processes. Elementary acts, such as Depict or Assert, are carried out by the graphics and text generators. In some situations, the generators might not be able to accomplish the requested acts. Thus, the planner has to modify its initial plan. In other cases, the generator might find out that there are better solutions. E.g., since the planner has no geometrical knowledge, it does not know when the graphics designer is able to incorporate information in a single picture. Thus, it requests the graphics designer to add a background for each depicted object. In case the graphics designer is able to convey the requested information in a single picture, this background has to be added only once. Consequently, the structure of the document can be simplified by factoring out the background.

When planning a document, one has to consider that information concerning decisions about the chosen modes of presentation can influence further decisions. As an example, we suppose that the system has to compare two objects. This can be done by describing the different values of a common attribute. At this time, the only restriction is that both descriptions should be realized in the same mode. Once the system has decided on the mode for the attribute value of the first object, this choice must be made available for describing the value of the second object. We tackle this problem by passing mode information during the planning process both from top to bottom and from bottom to top (cf. Fig. 2).

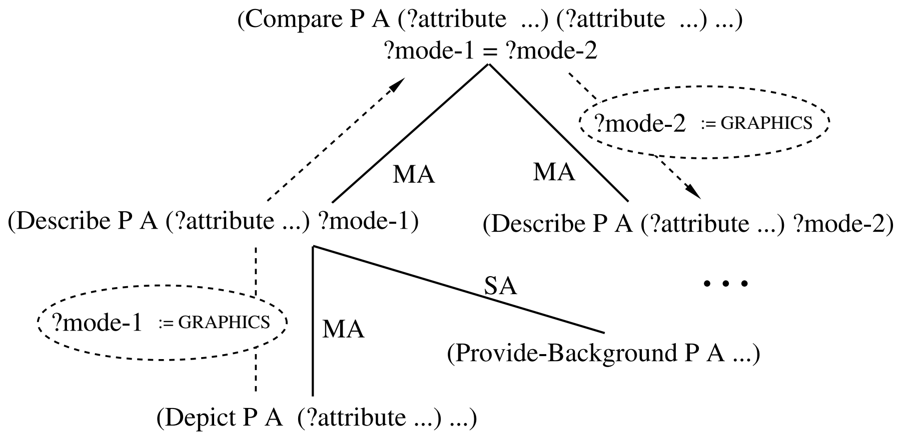


Fig. 2: Passing of information

Knowledge concerning the structure and the contents of the document at different stages is managed by a document handler. The document handler does not only serve as a database. It also transforms knowledge in a form readable for the component accessing it. E.g., to generate references to parts of a picture, geometrical data have to be transformed into spatial relations (cf. [André et al. 90]).

## 5 Building up Illustrated Documents: An Example

In the following, we will explain how the graphics and text generator are driven by the planner. Let's assume that the planner as the presenter has selected Strategy 1 to request the addressee to switch on the espresso-machine and sends the message (Request P A turn-1) to the text generator which starts with the utterance[7]. We presume that the addressee knows why the action should be carried out. Thus, we don't

[7] In this paper, we focus on content selection for text-picture combinations. Problems of content realization (e.g., lexical choice, image rendering) are not addressed here.

need to motivate him. To enable the action, Strategy 2 is used. According to Strategy 3, an action can be enabled by describing its result and its course. As mentioned earlier, a result of an action can be described by describing all changes happening to involved objects. In our case, only the orientation of knob-1 has changed. Since a change of orientation is considered as spatial (cf. section 3.1) and no mode restrictions have to be considered at this time, the planner decides to convey this information by showing the knob with its new orientation. Thus, the graphics generator receives the message (Depict P A (Orientation ori-1 knob-1) ?px ?pic) and designs a new picture pic-1 including the picture object p - 1 which refers to knob-1. The planner then checks whether the depicted knob can be identified. Since it's most probably not clear to the addressee which knob is shown, the planner decides to provide background information to resolve the ambiguity. In our example, this is accomplished by showing the spatial context (cf. Strategy 5). The graphics generator receives the message (Depict P A (Background em-1) ?pz pic-1) and modifies the current picture. As a consequence of modifications, initially visible objects might be occluded or no longer recognizable. After each change, the graphics generator, therefore, has to check whether the predicates Visible-P and Recognizable-P are still satisfied. In our example, the knob is no longer recognizable in detail. The graphics generator solves this problem by an inset. According to the subsidiary act in Strategy 3, the course of the turning action has to now be described. Since turning the knob involves motion, the trajectory of the knob is described by showing an arrow that indicates the direction of motion. Because the graphics generator is able to incorporate the arrow into the inset of pic-1, no further picture generaton is necessary. Some snapshots of the planning and generation processes are shown in Fig. 3.

Fig. 3: Plan-Based Synthesis of a Text-Picture Combination

## 6 Comparison to Related Research

Although there is broad interest in the automatic generation of multimodal explanations (cf. [Sullivan&Tyler 89] and [Arens et al. 89]), text and graphics are usually treated independently of each other. In this section, we focus on two other systems in which the integration of text and graphics is addressed: the system SAGE (cf. [Roth et al. 89] and [Roth&Mattis 90]) and the system COMET (cf. [Elhadad et al. 89] and [Feiner&McKeown 90]). The system SAGE automatically generates explanations of changes for quantitative modeling systems. While the graphics generator is based on the APT system (cf. [Mackinlay 88]), text planning is performed by using schemas (cf. [McKeown 85]). Roth and coworkers distinguish between data characteristics and the function of a presentation, but this information is primarily used to select graphical techniques. The COMET system automatically generates multimodal explanations of equipment maintenance and repair procedures. In contrast to SAGE, COMET uses schemas to produce a common content description for both generators. After content selection, the content description is annotated with mode indications according to the type of information to be conveyed. Compared to SAGE, this approach enables a tighter coordination between the two generators.

In contrast to SAGE and COMET, we use a plan-based approach for content selection as proposed in [Moore & Paris 89] for text generation. An important difference in our work concerns the flow of control. In our system, mode selection happens not after, but during content planning. As soon as the mode of presentation is determined, the information to be conveyed is sent to the generators. This method enables the planner to easily add information to be presented depending on the selected mode and the results of the generators. E.g., when generating a picture of a special object, the planner checks in a subsidiary act whether the picture is comprehensible, and adds pictorial or verbal elements if necessary.

## 7 Conclusions and Future Work

In this paper, we have argued that not only the generation of text, but also the synthesis of multimodal documents can be considered as a communicative act which aims to achieve certain goals. We have introduced presentation strategies to represent knowledge about presentation techniques. In order to decide between several presentation strategies, we have examined how the kind of information to be conveyed influences mode selection and which communicative functions single document parts play in text-picture combinations. For the realization of a system able to automatically generate illustrated documents, we have proposed a plan-based approach. The joint planning of text and pictures is regarded as a fundamental prerequisite for the coordination of different modes.

So far, we are concerned with the selection of the presentation mode for a given information content. Since in general one and the same information content can be expressed in a number of ways in a single mode, we also need criteria to rank the effectiveness of alternative graphical or textual presentations. E.g. when designing a picture, one has to ensure that its complexity does not exceed the addressee's perceptual and cognitive abilities.

Due to the planning process, each generated document part is associated with a certain communicative function specified in the presentation strategies. Thus, text and pictures are always related to each other. The communication process is, however, only successful if the addressee is able to derive the intended relations. In the future, we will therefore investigate how the relations between document parts can be effectively communicated. Research up to now has concentrated on textual relations (cf. [Scott&deSouza 90]). Text-picture researchers provide only a few hints. E.g., to convey a contrast relation between two pictures, the pictures should agree in form and style and be placed side by side. To provide evidence, realistic pictures should be chosen.

A prototype of the presentation planner and parts of the two generators have been implemented in Commonlisp on a MacIvory Lispsystem running Release 7.2. These components will be integrated into a complex tool for the knowledge-based presentation of information.

## References


[André et al. 90] E. **André**, G. **Herzog** and T. **Rist**. Natural Language Access to Visual Data: Dealing with Space and Movement. To appear in: F. Nef and M. Borillo (eds.): Semantics of Time, Space and Movement in Natural Language, Edition Hermes: Paris, 1990.

[Arens et al. 89] Y. **Arens**, S. **Feiner**, J. **Hollan**, B. **Neches**. Proceedings of the IJCAI-89 Workshop `A New Generation of Intelligent Interfaces´, Detroit, Michigan, 1989.

[Bandyopadhyay 90] S. **Bandyopadhyay**. Towards an Understanding of Coherence in Multimodal Discourse. Technical Memo DFKI-TM-90-01, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, 1990.

[Bieger&Glock 84] G.R. **Bieger** and M.D. **Glock**. The Information Content of Picture-Text Instructions. The Journal of Experimental Education 53(2), pp. 68-76, 1984.

[Bieger&Glock 86] G.R. **Bieger** and M.D. **Glock**. Comprehending Spatial and Contextual Information in Picture-Text Instructions. The Journal of Experimental Education 54(4), pp. 181-188, 1986.

[Brandt et al. 83] M. **Brandt**, W. **Koch**, W. **Motsch** and I. **Rosengren**. Der Einfluß der kommunikativen Strategie auf die Textstruktur - dargestellt am Beispiel des Geschäftsbriefes. In: I. Rosengren (Hrsg.), Sprache und Pragmatik, Lunder Symposium 1982. Almqvist & Wiksell: Stockholm, pp. 105-135, 1983.

[Cohen&Levesque 85] P.R. **Cohen** and H.J. **Levesque**. Speech Acts and Rationality. In: Proceedings of the 23rd Annual Meeting of the ACL, pp. 49-59, 1985.

[Van Dijk 80] T. A. **van Dijk**. Textwissenschaft. dtv: München, 1980.

[Elhadad et al. 89] M. **Elhadad**, D.D. **Seligmann**, S. **Feiner** and K.R. **McKeown**. A Common Intention Description Language for Interactive Multi-Media Systems. Proc. of the IJCAI-89 Workshop `A New Generation of Intelligent Interfaces´, Detroit, Michigan, 1989.

[Feiner&McKeown 90] S.K. **Feiner** and K.R. **McKeown**. Generating Coordinated Multimedia Explanations. In: 6th IEEE Conference on Artificial Intelligence Applications, Santa Barbara, CA, pp. 290-296, 1990.

[Grimes 75] J.E. **Grimes**. The Thread of Discourse. Mouton: The Hague, Paris, 1975.

[Hobbs 78] J. **Hobbs**. Why is a discourse coherent? Technical Report 176, SRI International, Menlo Park, CA, 1978.

[Hovy 88] E. H. **Hovy**. Approaches to the Planning of Coherent Text. Papers from the 4th International Workshop on Text Generation, Catalina Island, 1988.

[Hunter et al. 87] B. **Hunter**, A. **Crismore** and P.D. **Pearson**. Visual Displays in Basal Readers and Social Studies Textbooks. In: D.M. Willows and H. A. Houghton (eds.), The Psychology of Illustration, Basic Research, Vol. 2, Springer: New York, Berlin, Heidelberg, London, Paris, Tokyo, pp. 116-135, 1987.

[Kjorup 78] S. **Kjorup**. Pictorial Speech Acts. In: Erkenntnis 12, pp. 55-71, 1978.

[Levie 87] W.H. **Levie**. Research on Pictures: A Guide to the Literature. In: D.M. Willows and H. A. Houghton (eds.), The Psychology of Illustration, Basic Research, Vol. 1, Springer: New York, Berlin, Heidelberg, London, Paris, Tokyo, pp. 1-50, 1987.

[Levin et al. 87] J.R. **Levin**, G.J. **Anglin** and R. N. **Carney**. On Empirically Validating Functions of Pictures in Prose. In: D.M. Willows and H. A. Houghton (eds.), The Psychology of Illustration, Basic Research, Vol. 1, Springer: New York, Berlin, Heidelberg, London, Paris, Tokyo, pp. 51-85, 1987.

[Mackinlay 88] J. **Mackinlay**. Search Architecture for the Automatic Design of Graphical Presentations. In: J. Sullivan and S. Tyler (eds.), Architectures for Intelligent Interfaces: Elements and Prototypes, Addison-Wesley, 1989.

[Mann&Thompson 87] W.C. **Mann** and S.A. **Thompson**. Rhetorical Structure Theory: Description and Construction of Text Structures. In: G. Kempen (eds.), Natural Language Generation: New Results in Artificial Intelligence, Psychology, and Linguistics, Nijhoff: Dordrecht, Boston, Lancaster, pp. 85-95, 1987.

[McKeown 85] K.R. **McKeown**. Text Generation. Cambridge University Press: London, 1985.

[Molitor et al. 89] S. **Molitor**, S.-P. **Ballstaedt** and H. **Mandl**. Problems in Knowledge Acquisition from Text and Pictures. In: H. Mandl and J.R. Levin (eds.): Knowledge Acquisition from Text and Pictures. North Holland: Amsterdam, New York, Oxford, Tokyo, pp. 3-35, 1989.

[Moore&Paris 89] J.D. **Moore** and C.L. **Paris**. Planning Text for Advisory Dialogues. In: Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, 1989.

[Moore&Swartout 89] J.D. **Moore** and W.R. **Swartout**. A Reactive Approach to Explanation. In: Proceedings of the 11th International Joint Conference on Artificial Intelligence, 1989.

[Muckenhaupt 86] M. **Muckenhaupt**. Text und Bild. Gunter Narr: Tübingen, 1986.

[Roth et al. 89] S.F. **Roth**, J. **Mattis** and X. **Mesnard**. Graphics and Natural Language as Components of Automatic Explanation. In: J. Sullivan and S. Tyler (eds.), Architectures for Intelligent Interfaces: Elements and Prototypes, Addison-Wesley, 1989.

[Roth&Mattis 90] S.F. **Roth** and J. **Mattis**. Data Characterization for Intelligent Graphics Presentation. In: Proc. of CHI 90, 1990.

[Scott&deSouza 90] D.R. **Scott**, C.S. **deSouza**. Getting the Message Across in RST-based Text Generation. In: R. Dale, C. Mellish and M. Zock (eds.), Current Research in Natural Language Generation, Academic Press, 1990.

[Searle 69] J.R. **Searle**. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press: Cambridge, MA, 1969.

[Smith&Smith 66] K.U. **Smith** and M.F. **Smith**. Cybernetic Principles of Learning and Educational Design. Holt, Rinehart & Winston: New York, 1966.

[Strothotte&Schmid 90] T. **Strothotte** and C. **Schmid**. Semiformale Darstellungen in wissensbasierten Systemen. In: K. Kansy and P. Wißkirchen (eds.), Graphik und KI, GI-Fachgespräch, Springer: Berlin, Heidelberg, pp. 1-9, 1990.

[Sullivan&Tyler 89] J.W. **Sullivan** and S.W. **Tyler** (eds.). Architectures for Intelligent User Interfaces: Elements and Prototypes. Addison-Wesley, 1989.

[Wahlster et al. 89] W. **Wahlster**, E. **André**, M. **Hecking** and T. **Rist**. WIP: Knowledge-based Presentation of Information. Report WIP-1, Deutsches Forschungszentrum für Künstliche Intelligenz, Saarbrücken, 1989.

Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

DFKI
-Bibliothek-
PF 2080
67608 Kaiserslautern
FRG

# DFKI Publikationen

Die folgenden DFKI Veröffentlichungen sowie die aktuelle Liste von allen bisher erschienenen Publikationen können von der oben angegebenen Adresse oder per anonymem ftp von ftp.dfki.uni-kl.de (131.246.241.100) unter pub/Publications bezogen werden.
Die Berichte werden, wenn nicht anders gekenn-zeichnet, kostenlos abgegeben.

# DFKI Publications

The following DFKI publications or the list of all published papers so far are obtainable from the above address or per anonymous ftp from ftp.dfki.uni-kl.de (131.246.241.100) under pub/Publications.
The reports are distributed free of charge except if otherwise indicated.

**DFKI Research Reports**

**RR-92-43**
*Christoph Klauck, Jakob Mauss:* A Heuristic driven Parser for Attributed Node Labeled Graph Grammars and its Application to Feature Recognition in CIM
17 pages

**RR-92-44**
*Thomas Rist, Elisabeth André:* Incorporating Graphics Design and Realization into the Multimodal Presentation System WIP
15 pages

**RR-92-45**
*Elisabeth André, Thomas Rist:* The Design of Illustrated Documents as a Planning Task
21 pages

**RR-92-46**
*Elisabeth André, Wolfgang Finkler, Winfried Graf, Thomas Rist, Anne Schauder, Wolfgang Wahlster:* WIP: The Automatic Synthesis of Multimodal Presentations
19 pages

**RR-92-47**
*Frank Bomarius:* A Multi-Agent Approach towards Modeling Urban Traffic Scenarios
24 pages

**RR-92-48**
*Bernhard Nebel, Jana Koehler:*
Plan Modifications versus Plan Generation:
A Complexity-Theoretic Perspective
15 pages

**RR-92-49**
*Christoph Klauck, Ralf Legleitner, Ansgar Bernardi:*
Heuristic Classification for Automated CAPP
15 pages

**RR-92-50**
*Stephan Busemann:*
Generierung natürlicher Sprache
61 Seiten

**RR-92-51**
*Hans-Jürgen Bürckert, Werner Nutt:*
On Abduction and Answer Generation through Constrained Resolution
20 pages

**RR-92-52**
*Mathias Bauer, Susanne Biundo, Dietmar Dengler, Jana Koehler, Gabriele Paul:* PHI - A Logic-Based Tool for Intelligent Help Systems
14 pages

**RR-92-53**
*Werner Stephan, Susanne Biundo:*
A New Logical Framework for Deductive Planning
15 pages

**RR-92-54**
*Harold Boley:* A Direkt Semantic Characterization of RELFUN
30 pages

**RR-92-55**
*John Nerbonne, Joachim Laubsch, Abdel Kader Diagne, Stephan Oepen:* Natural Language Semantics and Compiler Technology
17 pages

**RR-92-56**
*Armin Laux:* Integrating a Modal Logic of Knowledge into Terminological Logics
34 pages

**RR-92-58**
*Franz Baader, Bernhard Hollunder:*
How to Prefer More Specific Defaults in Terminological Default Logic
31 pages

**RR-92-59**
*Karl Schlechta and David Makinson:* On Principles and Problems of Defeasible Inheritance
13 pages

**RR-92-60**
*Karl Schlechta:* Defaults, Preorder Semantics and Circumscription
19 pages

**RR-93-02**
*Wolfgang Wahlster, Elisabeth André, Wolfgang Finkler, Hans-Jürgen Profitlich, Thomas Rist:*
Plan-based Integration of Natural Language and Graphics Generation
50 pages

**RR-93-03**
*Franz Baader, Berhard Hollunder, Bernhard Nebel, Hans-Jürgen Profitlich, Enrico Franconi:*
An Empirical Analysis of Optimization Techniques for Terminological Representation Systems
28 pages

**RR-93-04**
*Christoph Klauck, Johannes Schwagereit:*
GGD: Graph Grammar Developer for features in CAD/CAM
13 pages

**RR-93-05**
*Franz Baader, Klaus Schulz:* Combination Tech-niques and Decision Problems for Disunification
29 pages

**RR-93-06**
*Hans-Jürgen Bürckert, Bernhard Hollunder, Armin Laux:*
On Skolemization in Constrained Logics
40 pages

**RR-93-07**
*Hans-Jürgen Bürckert, Bernhard Hollunder, Armin Laux:*
Concept Logics with Function Symbols
36 pages

**RR-93-08**
*Harold Boley, Philipp Hanschke, Knut Hinkelmann, Manfred Meyer:* COLAB: A Hybrid Knowledge Representation and Compilation Laboratory
64 pages

**RR-93-09**
*Philipp Hanschke, Jörg Würtz:*
Satisfiability of the Smallest Binary Program
8 Seiten

**RR-93-10**
*Martin Buchheit, Francesco M. Donini, Andrea Schaerf:*
Decidable Reasoning in Terminological Knowledge Representation Systems
35 pages

**RR-93-11**
*Bernhard Nebel, Hans-Juergen Buerckert:*
Reasoning about Temporal Relations:
A Maximal Tractable Subclass of Allen's Interval Algebra
28 pages

**RR-93-12**
*Pierre Sablayrolles:* A Two-Level Semantics for French Expressions of Motion
51 pages

**RR-93-13**
*Franz Baader, Karl Schlechta:*
A Semantics for Open Normal Defaults via a Modified Preferential Approach
25 pages

**RR-93-14**
*Joachim Niehren, Andreas Podelski,Ralf Treinen:*
Equational and Membership Constraints for Infinite Trees
33 pages

**RR-93-15**
*Frank Berger, Thomas Fehrle, Kristof Klöckner, Volker Schölles, Markus A. Thies, Wolfgang Wahlster:* PLUS - Plan-based User Support
Final Project Report
33 pages

**RR-93-16**
*Gert Smolka, Martin Henz, Jörg Würtz:* Object-Oriented Concurrent Constraint Programming in Oz
17 pages

**RR-93-17**
*Rolf Backofen:*
Regular Path Expressions in Feature Logic
37 pages

**RR-93-18**
*Klaus Schild:* Terminological Cycles and the Propositional μ-Calculus
32 pages

**RR-93-20**
*Franz Baader, Bernhard Hollunder:*
Embedding Defaults into Terminological Knowledge Representation Formalisms
34 pages

**RR-93-22**
*Manfred Meyer, Jörg Müller:*
Weak Looking-Ahead and its Application in Computer-Aided Process Planning
17 pages

**RR-93-23**
*Andreas Dengel, Ottmar Lutzy:*
Comparative Study of Connectionist Simulators
20 pages

**RR-93-24**
*Rainer Hoch, Andreas Dengel:*
Document Highlighting —
Message Classification in Printed Business Letters
17 pages

**RR-93-25**
*Klaus Fischer, Norbert Kuhn:* A DAI Approach to Modeling the Transportation Domain
93 pages

**RR-93-26**
*Jörg P. Müller, Markus Pischel:* The Agent Architecture InteRRaP: Concept and Application
99 pages

**RR-93-27**
*Hans-Ulrich Krieger:*
Derivation Without Lexical Rules
33 pages

**RR-93-28**
*Hans-Ulrich Krieger, John Nerbonne, Hannes Pirker:* Feature-Based Allomorphy
8 pages

**RR-93-29**
*Armin Laux:* Representing Belief in Multi-Agent Worlds viaTerminological Logics
35 pages

**RR-93-33**
*Bernhard Nebel, Jana Koehler:*
Plan Reuse versus Plan Generation: A Theoretical and Empirical Analysis
33 pages

**RR-93-34**
Wolfgang Wahlster:
Verbmobil Translation of Face-To-Face Dialogs
10 pages

**RR-93-35**
*Harold Boley, François Bry, Ulrich Geske (Eds.):* Neuere Entwicklungen der deklarativen KI-Programmierung — *Proceedings*
150 Seiten
**Note:** This document is available only for a nominal charge of 25 DM (or 15 US-$).

**RR-93-36**
*Michael M. Richter, Bernd Bachmann, Ansgar Bernardi, Christoph Klauck, Ralf Legleitner, Gabriele Schmidt:* Von IDA bis IMCOD: Expertensysteme im CIM-Umfeld
13 Seiten

**RR-93-38**
*Stephan Baumann:* Document Recognition of Printed Scores and Transformation into MIDI
24 pages

**RR-93-40**
*Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, Werner Nutt, Andrea Schaerf:*
Queries, Rules and Definitions as Epistemic Statements in Concept Languages
23 pages

**RR-93-41**
*Winfried H. Graf:* LAYLAB: A Constraint-Based Layout Manager for Multimedia Presentations
9 pages

**RR-93-42**
*Hubert Comon, Ralf Treinen:*
The First-Order Theory of Lexicographic Path Orderings is Undecidable
9 pages

**RR-93-45**
*Rainer Hoch:* On Virtual Partitioning of Large Dictionaries for Contextual Post-Processing to Improve Character Recognition
21 pages

## DFKI Technical Memos

**TM-91-14**
*Rainer Bleisinger, Rainer Hoch, Andreas Dengel:*
ODA-based modeling for document analysis
14 pages

**TM-91-15**
*Stefan Busemann:* Prototypical Concept Formation An Alternative Approach to Knowledge Representation
28 pages

**TM-92-01**
*Lijuan Zhang:* Entwurf und Implementierung eines Compilers zur Transformation von Werkstückrepräsentationen
34 Seiten

**TM-92-02**
*Achim Schupeta:* Organizing Communication and Introspection in a Multi-Agent Blocksworld
32 pages

**TM-92-03**
*Mona Singh:*
A Cognitiv Analysis of Event Structure
21 pages

**TM-92-04**
*Jürgen Müller, Jörg Müller, Markus Pischel, Ralf Scheidhauer:*
On the Representation of Temporal Knowledge
61 pages

**TM-92-05**
*Franz Schmalhofer, Christoph Globig, Jörg Thoben:*
The refitting of plans by a human expert
10 pages

**TM-92-06**
*Otto Kühn, Franz Schmalhofer:* Hierarchical skeletal plan refinement: Task- and inference structures
14 pages

**TM-92-08**
*Anne Kilger:* Realization of Tree Adjoining Grammars with Unification
27 pages

**TM-93-01**
*Otto Kühn, Andreas Birk:* Reconstructive Integrated Explanation of Lathe Production Plans
20 pages

**TM-93-02**
*Pierre Sablayrolles, Achim Schupeta:*
Conlfict Resolving Negotiation for COoperative Schedule Management
21 pages

**TM-93-03**
*Harold Boley, Ulrich Buhrmann, Christof Kremer:*
Konzeption einer deklarativen Wissensbasis über recyclingrelevante Materialien
11 pages

## DFKI Documents

**D-92-19**
*Stefan Dittrich, Rainer Hoch:* Automatische, Deskriptor-basierte Unterstützung der Dokument-analyse zur Fokussierung und Klassifizierung von Geschäftsbriefen
107 Seiten

**D-92-21**
*Anne Schauder:* Incremental Syntactic Generation of Natural Language with Tree Adjoining Grammars
57 pages

**D-92-22**
*Werner Stein:* Indexing Principles for Relational Languages Applied to PROLOG Code Generation
80 pages

**D-92-23**
*Michael Herfert:* Parsen und Generieren der Prolog-artigen Syntax von RELFUN
51 Seiten

**D-92-24**
*Jürgen Müller, Donald Steiner (Hrsg.):*
Kooperierende Agenten
78 Seiten

**D-92-25**
*Martin Buchheit:* Klassische Kommunikations- und Koordinationsmodelle
31 Seiten

**D-92-26**
*Enno Tolzmann:*
Realisierung eines Werkzeugauswahlmoduls mit Hilfe des Constraint-Systems CONTAX
28 Seiten

**D-92-27**
*Martin Harm, Knut Hinkelmann, Thomas Labisch:*
Integrating Top-down and Bottom-up Reasoning in COLAB
40 pages

**D-92-28**
*Klaus-Peter Gores, Rainer Bleisinger:* Ein Modell zur Repräsentation von Nachrichtentypen
56 Seiten

**D-93-01**
*Philipp Hanschke, Thom Frühwirth:* Terminological Reasoning with Constraint Handling Rules
12 pages

**D-93-02**
*Gabriele Schmidt, Frank Peters,*
*Gernod Laufkötter:* User Manual of COKAM+
23 pages

**D-93-03**
*Stephan Busemann, Karin Harbusch(Eds.):*
DFKI Workshop on Natural Language Systems: Reusability and Modularity - Proceedings
74 pages

**D-93-04**
DFKI Wissenschaftlich-Technischer Jahresbericht 1992
194 Seiten

**D-93-05**
*Elisabeth André, Winfried Graf, Jochen Heinsohn, Bernhard Nebel, Hans-Jürgen Profitlich, Thomas Rist, Wolfgang Wahlster:*
PPP: Personalized Plan-Based Presenter
70 pages

**D-93-06**
*Jürgen Müller (Hrsg.):*
Beiträge zum Gründungsworkshop der Fachgruppe Verteilte Künstliche Intelligenz Saarbrücken 29.-30. April 1993
235 Seiten
**Note:** This document is available only for a nominal charge of 25 DM (or 15 US-$).

**D-93-07**
*Klaus-Peter Gores, Rainer Bleisinger:*
Ein erwartungsgesteuerter Koordinator zur partiellen Textanalyse
53 Seiten

**D-93-08**
*Thomas Kieninger, Rainer Hoch:* Ein Generator mit Anfragesystem für strukturierte Wörterbücher zur Unterstützung von Texterkennung und Textanalyse
125 Seiten

**D-93-09**
*Hans-Ulrich Krieger, Ulrich Schäfer:*
TDL ExtraLight User's Guide
35 pages

**D-93-10**
*Elizabeth Hinkelman, Markus Vonerden,Christoph Jung:*
Natural Language Software Registry
(Second Edition)
174 pages

**D-93-11**
*Knut Hinkelmann, Armin Laux (Eds.):*
DFKI Workshop on Knowledge Representation Techniques — Proceedings
88 pages

**D-93-12**
*Harold Boley, Klaus Elsbernd, Michael Herfert, Michael Sintek, Werner Stein:*
RELFUN Guide: Programming with Relations and Functions Made Easy
86 pages

**D-93-14**
*Manfred Meyer (Ed.):* Constraint Processing – Proceedings of the International Workshop at CSAM'93, July 20-21, 1993
264 pages
**Note:** This document is available only for a nominal charge of 25 DM (or 15 US-$).