

Knowledge Representation for Automatic Technical Text Comprehension

Grigoriadou M.*¹, Blitsas P.², Marouli A.¹

1. Department of Informatics and Telecommunications

2. Interdisciplinary Program of Graduate Studies in Basic and Applied Cognitive Science
National and Kapodistrian University of Athens, Greece,
Panepistimioupolis 15771, Ilissia, Athens, Greece

Abstract -- In the present work, a system of computational ontology is presented, for representing the knowledge of technical text in the context of Computer Networks domain, according to the Denhière-Baudet model for technical text comprehension. The semantic relations among entities and events of a technical text can be expressed by structures, as microstructure and macrostructure, of this system. In this paper, the explicit and implicit knowledge representation of two extracts of a technical text is provided, as well as, the teleological representation of the functional system operations, depicted in these extracts. The presented methodology can support automatic reasoning, through the combination of teleology and causality, which can be used in automatic knowledge extraction from a technical text body.

Index terms -- state, event, microstructure, macrostructure, taxonomy, partonomy, functional system, functional object analysis, teleology, inference.

I. INTRODUCTION

Today, a lot of effort is being put for creating programs and algorithms to simulate psycholinguistic theories with regard to human processes of memory and text comprehension, as the construction-integration model [9], the theory of latent semantic analysis of knowledge representation [8], and the Denhière & Baudet technical text comprehension model [1]. The text comprehension models use cognitive representations which focus in deeper levels of understanding, such as conceptual conclusions based on knowledge (knowledge-based inferences), inductive reductions (pragmatics) or world knowledge combined with the surface levels of understanding which involve lexical processing, syntactic analysis and text interpretation [3], [11]. Within the framework of the present work, the Denhière & Baudet technical text comprehension model and two indicative examples of knowledge representation for automatic technical text comprehension based on the explicit and implicit content of two extracts of technical text, are presented. The knowledge representation of these extracts content can reveal the importance of the teleology to knowledge extraction through the automatic inferences.

* Corresponding author. Tel.:+30-210-7275205
E-mail address: gregor@di.uoa.gr (M. Grigoriadou)

II. DENHIÈRE – BAUDET MODEL OF TECHNICAL TEXT COMPREHENSION

According to the text comprehension model of Deniere and Baudet [1] the individual who reads a text manufactures progressively the microstructure of its representation, namely states, events and actions of the world that is described in text, as well as, temporal and causal relations which connect these structures. The term ‘state’ is static and it describes a situation where no change happens during time. The term ‘event’ describes an action that causes changes but is not evoked by a person. An event can be accidental or can be caused by non-human action e.g. a machine. An ‘action’ causes changes and is evoked by a person. The above fundamental cognitive categories are determined by the characteristics of an individual’s cognitive system. The text should support the construction of the following representations:

- Microstructure, which includes a description of system entities based on temporal/causal relations that connect them, a description of sequence of events executed on these entities, and the changes caused by these events to the states of these entities.
- Macrostructures, which includes the teleological and hierarchical structure of goal and sub-goals of various operations of the system.

III. TECHNICAL TEXT – FUNCTIONAL SYSTEM

A technical text depicts a technical system, which contains a set of associated entities, fixed by hierarchical relations of whole/part type. A technical system, organized as a tree of goal/subgoals, is called a “Functional System”. The “construction” of a text, which allows a precise description of a technical system and facilitates a reader in manufacturing a macrostructure, is called “Technical Object Analysis of a Functional System”. A technical text is supposed to provide descriptions which include *relational structure*, which is a set of entities characterized by values, received by their characteristics, and influenced by the determination of one or more static relations among these entities, *transformation*, which is the modification of natural tendency of progressive system states, which either remain the same or change, and *teleology*, which is the initial system state that is determined by the present entities, their relations and the values of their attributes. This initial system state changes in order to reach a final state by having a determined schematization that constitutes the predetermined system goal. Each modification has the operation of achieving this goal.

IV. TWO INDICATIVE EXAMPLES OF KNOWLEDGE REPRESENTATION FOR AUTOMATIC TECHNICAL TEXT COMPREHENSION BASED ON FUNCTIONAL SYSTEM CAUSALITY AND TELEOLOGY

A. *Microstructure Construction*

At the microstructure level, there is a description of units which constitute the system, as well as, the causal relations which connect them. Initial stage for realizing microstructure of a text that belongs in a technical domain is creating taxonomy, and paronomy of the concepts which are presented in the text. The taxonomy concerns the entities that “participate” in the technical text comprehension, particularly in Textbase level [8]. This ontology constitutes a tree-like structure, “is_a” type, which analyzes the taxonomical relations among the entities. The paronomy is a tree-like structure too, “part_of” or “has” type, which expresses whole/part relations among the entities of the microcosm. The basic characteristic of the microstructure is the fact that it has to depict the explicit knowledge of the text, as well as, the implicit knowledge, which is contained in the same text and is activated during the reader’s reading and comprehending.

The first extract, was taken from the book “Computer Science: an overview” [2]. It refers to combination of bus topology networks towards forming wide area computer networks:

The repeater is little more than a device that connects two buses to form a single long bus. The repeater simply passes signals back and forth between the two original buses (usually with some form of amplification) without considering the meaning of the signals. A bridge is similar to, but more complex than, a repeater. Like a repeater, it connects two buses, but it does not necessarily pass all messages across the connection. Instead, it looks at the destination address that accompanies each message and forwards a message across the connection only when that message is destined for a computer on the other side.

Explicit knowledge representation of the extract

Taxonomy

Repeater is a device	is_a(repeater, device).	Bus2 is an instance of bus	instance_of(bus2, bus).
Bus1 is an instance of bus	instance_of(bus1, bus).	Forward is send	is_a(forward, send).
Amplify is increase	is_a(amplify, increase).	Pass is send	is_a(pass, send).

Events

Repeater connects two buses	connect_without_control(repeater, bus1, bus2).
Two buses forms a single long bus	form(bus1, bus2, long_bus).
Repeater passes signals back and forth between 2 buses	pass(repeater, signal, back_forth, no_control).
Repeater amplifies signals	amplify(repeater, message).
Bridge connects two buses	connect_with_control(bridge, bus1, bus2).

Bridges passes the message only when it is destined for a computer on the other side
 Bridge looks at the destination address
 There is no consideration of signals meaning
 Message is accompanied by a destination address

```
pass(bridge, signal, back_forth, control).
look(bridge, destination_address).
no_control(signal_meaning).
has(message, destination_address).
```

Implicit knowledge representation of the extract

Partonomy

Node is a part of a bus.	part_of(node, bus).
Lan is a part of wan.	part_of(lan, wan).
Content is a part of signal.	part_of(content, signal).
Content has a meaning.	has(content, meaning).
Bus has CSMA/CD.	has(bus, csma_cd_protocol).
Bus has address.	has(bus, address)

Taxonomy

Bus is a lan.	Is_a(bus, lan).
Lan is a network.	is_a(lan, network).
Message is data.	is_a(message, data).
Data is a signal.	is_a(data, signal).
Bus is a topology.	is_a(bus, topology).
Bridge is a device	Is_a(bridge, device).

Transformation-Sequence of Events

In the case of a repeater

Source node sends the message to the source bus
 Message is transmitted to the bus of the other side through repeater
 If the destination bus is not occupied
 Message is delivered to destination node,
 Else it waits until destination bus is not occupied.

```
message_transmission_without_control:-
send(source_node, message, source_bus),
send(source_bus, message, repeater),
amplify(repeater, message),
send(repeater, message, destination_bus)
send(destination_bus, message, destination_node).
```

In the case of a bridge:

Bridge detects the destination address of the message
 If message refers to bus2 then if bus2 is not occupied
 bridge sends the message to the bus2
 And Message is delivered to destination node2 in bus2.
 Else it waits until bus2 is not occupied.
 Else bridge doesn't send it to bus2
 And if bus2 is not occupied
 Message is delivered to destination node2 in bus1.
 Else it waits until bus1 is not occupied.

```
message_transmission_with_control:-
send(source_node,message,bus1),
send(bus1,message,bridge),
detect(bridge, destination_address),
control(bridge,destination_address,address,1) *,
send(bridge,message,bus2),
send(bus2,message,destination_node)**;
send(bus1,message,destination_node)**.
* control(bridge, destination_address,address,C_value):-
destination_address=address, C_value is 1.
**send(instance_of(X, bus) ,destination_node):-
bus_situation(instance_of(X, bus)) = 0; wait.
```

The second extract, was taken from the book “Computer Science: An overview” [2], too. It refers to computer main memory and its operation.

A computer's main memory is organized in manageable units called cells, with a typical cell size being eight bits.. The left end of this row is called the high-order end, and the right end is called the low-order end. The last bit at the high-order end is called the high-order bit or the most significant bit (MSB) in reference to the fact that if the contents of the cell were interpreted as representing a numeric value, this bit would be the most significant digit in the number. Similarly the bit at the right end is referred to as the low-order bit or the least significant bit (LSB). To identify individual cells in a computer's main memory, each cell is assigned a unique name, called its address. ...

the circuitry that actually holds the bits is combined with the circuitry required to allow other circuits to store and retrieve data from the memory cells. In this way other circuits can get data from the memory by electronically asking for the contents of a certain address (called a read operation), or they can record information in the memory by requesting that a certain bit pattern be placed in the cell at a particular address (called a write operation).

Explicit knowledge representation of the extract

Taxonomy

Main Memory. is a storage device.	is_a(main_memory, storage_device).
The MSB is a high order bit.	is_a(msb, high_order_bit)
The LSB is a low order bit.	is_a(lsb, low_order_bit).
Main memory is a random access memory.	is_a(main_memory, ram)

Partonomy

Main memory is a part of computer memory.	part_of (main_memory, computer_memory).
Cell is a part of main memory.	part_of(cell, main_memory).
Circuit is a part of cell.	part_of(circuit, cell).
Main memory has a number of cells	has(main_memory , number_of_cells)
A cell has content.	has(cell, content).
A cell has MSB.	has(cell, msb).
A cell has LSB.	has(cell, lsb).
A cell has an address.	has(cell, address).
A byte has 8 bits.	has(byte, 8_bits).

Implicit knowledge representation of the extract

Transformation-Sequence of Events

Operation of reading data from the main memory
 Request for reading data from the cell with address.
 Detection of the content of the cell with address.
 Loading the content to the destination register.
 The destination register has the new content.

read-operation (main_memory, address):-
 request_data(read, main_memory, address),
 detect(address, content),
 load (content, destination_register),
 destination_register(**new_content**).

Operation of writing data to the main memory
 Request for writing data to the cell with address.
 Storage of data to the cell with address.
 The cell with address has a new content.

write-operation (main_memory, address):-
 request(write, main_memory, address),
 store(address, content),
 cell(**new_content**).

B. Macrostructure Construction

At the macrostructure level, for the first text there is a description, the main goal of the devices used for combining local networks, as well as, subgoals that result from these goals. The description of the latter goal/subgoals is expressed as a goal and subgoals arboreal structure. Predication rules to express the macrostructure for representing operations of combining two buses with and without control are following (schematically shown in figure 1):

Data Transferring between buses without control has as subgoal Data Amplification.	transfer_without_control(Device , data, bus1,bus2):- amplify(Device , Data).
Data Transferring between two buses with control has as subgoal Data Control.	transfer_with_control(Device , data, bus1,bus2):- control(Device , data_address).
Data Amplification has as subgoal Connecting two buses without control.	amplify(Device , Data):- connect_without_control(Device , bus1,bus2)
Data Amplification has as subgoal signal increase.	amplify(Device , Data):- increase(Signal).
Connecting two buses without control has as subgoal Data transferring within a bus, Device is a Repeater.	connect_without_control(Device , bus1,bus2):- transfer_in_lan(data,node1, node2), Device is repeater.
Data control has as subgoal Comparison of data address with bus2 address.	control(Device , data_address):- compare(Device , data_address,bus2_address).
Comparison of data address with bus2 address has as subgoal Connecting two buses with control.	compare(Device , data_address,bus2_address), connect_with_control(Device , bus1,bus2)
Connecting two buses with control has as subgoal, Data transferring within a bus, Connecting Device is a Bridge.	connect_with_control(Device , bus1,bus2):- transfer_in_lan(data,node1, node2), Device is bridge.

The same procedure has been followed for the second text. (Shown, schematically, in figure 2)

Reading data from the main memory has as subgoal Requesting for reading data from the cell	Read-operation (Device , address):- Request(read, Device , address).
Requesting for reading data from the cell with address has as subgoal Detection of the content of the cell	Request(read, Device ,address):-Detect(Device , address, data), Device is main_memory
Detection of the content of the cell has as subgoal Loading the content to the destination register	Detect(Device , address):- Load (Device , data, destination_register).
Writing data to the main memory has as subgoal Requesting for writing data to the cell	Write-operation (Device ,address):- Request(write, Device ,address).
Requesting for writing data to the cell has as subgoal Storage of data to the cell	Request(write, Device ,address):- Store(address, data), Device is main_memory.

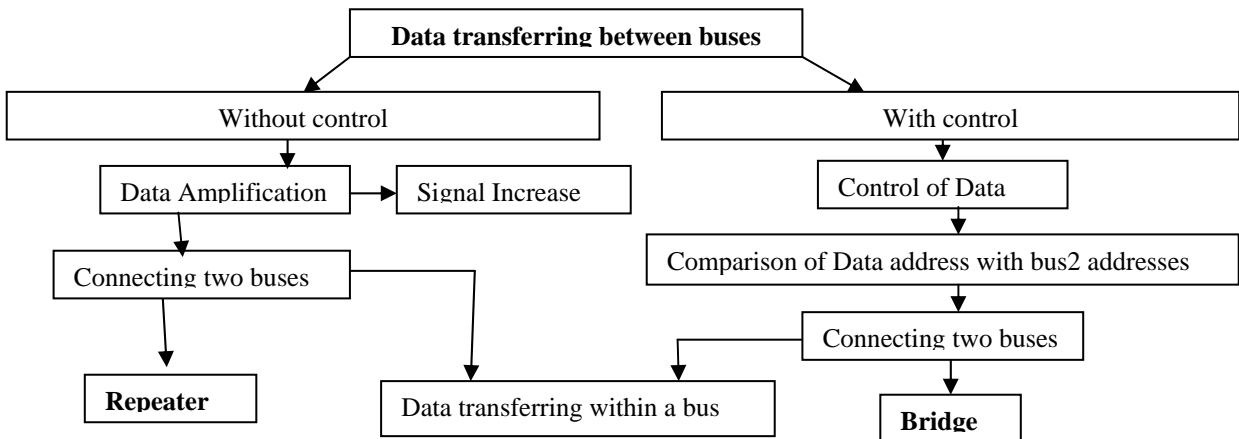


Figure1. Macrostructure Teleology of the first example

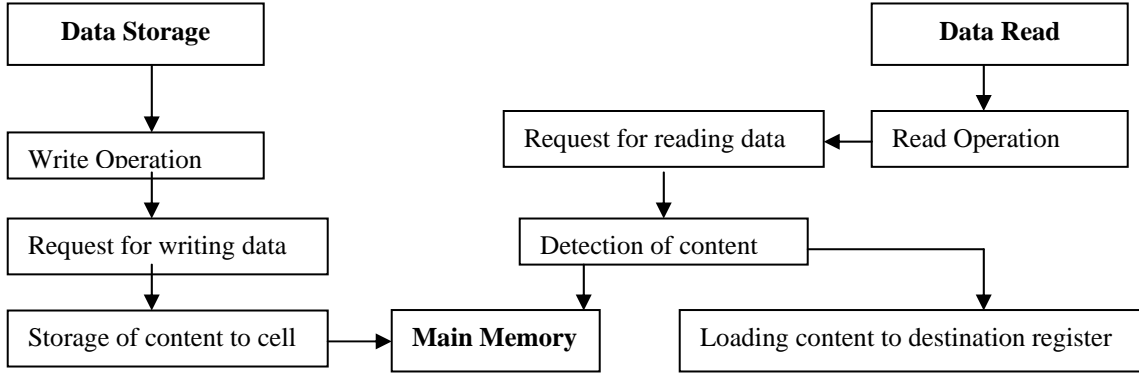


Figure2. Macrostructure Teleology of the second example

V. INFERENCE RULES IMPLEMENTATION

After the analysis in Microstructure and Macrostructure level, which was implemented in the previous two examples, predication rules [7] are presented, which could be used for knowledge extraction, through **automatic reasoning**.

which_is_Device_to_combine_2_buses_with_control(Device):-transfer_with_control(Device, data,bus1,bus2).

which_is_Device_for_storing_data_in_a_computer (Device):-write_operation(Device, address).

These rules can illustrate a way of extracting information through the teleological structure.

The first query returns the “**bridge**”, and the second query returns the value “**main_memory**”.

VI. DISCUSSION

The main purpose of this work was to illustrate the knowledge representation upon a technical text based on the model of Denhière-Baudet for technical text comprehension. The fact that the microstructure is not sufficient to describe a technical system, but the macrostructure is necessary, as well, is revealed. The complete analysis in microstructure and macrostructure level helps can support automatic reasoning. Automatic reasoning, in turn, leads to automatic comprehension in form of knowledge extraction through these two levels. The present research could be extended toward an automatic short free-text response assessment. So far, the most of the systems for computer assisted assessment of free-text responses are based on the *Latent Semantic Analysis Theory* [11]. As mentioned in [13], the main disadvantage of these systems is the fact that they are not incremental, because they are based on vector-oriented analysis. The present methodology could give a boost to a knowledge-based assessment system, with an ability of ex-

tending the knowledge base by adding new conceptual rules, expressing more causal, special and temporal relations, as well as, more goals and subgoals, which can describe new operations of a functional system, which in turn describes the content of a new technical text body.

VII. FUTURE PLANS

A Semantic Space, which is a mathematical representation of a large body of text, and every term, every text, and every novel combination of terms has a high dimensional vector representation, according to recent research, is different between novices and experts [9], [13]. The implicit knowledge must not be defined only by the experts' knowledge but the reader's prior knowledge should be taken under consideration, as well. Our short-term plans include a research on the semantic networks of novices and experts which can provide explanatory data about the differences in technical text understanding, based on the prior knowledge in Informatics science and especially in computer network domain [10].

ACKNOWLEDGEMENTS

The work, described in this paper, was partially supported by a grant from the project "Language Engineering Tools in Learning Environment: Application, Research, Innovation (Let's LEARN)", which is funded by The General Secretariat for Research and Technology (Contract GSRT 092-g).

REFERENCES

- [1] Baudet S, & Denhière G, 1992 Lecture, compréhension de texte et science cognitive. Presses Universitaires de France, Paris.
- [2] Brookshear G. , 2006, Computer Science : An Overview, Pearson Addison Wesley., Ninth Edition.
- [3] Goldman, S. R., Golden, R. M., Thomas, C., Oney, B., Macleod, S., & Lauterbach, M., 2003. Applications of Text Processing Models to Diagnostic Assessment of Reading. Thirteenth Annual Meeting of the Society for Text and Discourse. June 26-28, 2003. Madrid, Spain.
- [4] Graesser A. & Tipping P., 1999, Chapter 24: Understanding texts, In W. Bechtel & G. Graham (Eds.), A Companion to Cognitive Science, Malden MA: Blackwell.
- [5] Gruber T., 1993 Toward Principles for the Design of Ontologies Used for Knowledge Sharing Technical Report KSL 93-04, Stanford University, Knowledge Systems Laboratory. Revision.
- [6] Kanidis E. & Grigoriadou M, 2007, Reading about Computer Cache Memory: The Effects of Text Structure in Science Learning, In Proc. of 2th European Conf. on Cognitive Science (EuroCogSci07 2007).
- [7] Kintsch, W., 2001 Predication. *Cognitive Science* 25, 173-202
- [8] Kintsch, W., 1992, A cognitive architecture for comprehension. In H. L. Pick, P. van den Broek, & D. C. Knill (Eds.), The study of cognition: Conceptual and methodological issues (p. 143-164). Washington, DC: American Psychological Association.
- [9] Landauer T. K., & Dumais, S. T. 1997, A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2) , 211-240.
- [10] Lemaire B, & Denhière G et al, 2006, A computational model for simulating text comprehension, Behavior Research methods.
- [11] Liu H., & Singh P., 2004. Commonsense reasoning in and over natural language. In Proc. of 8th Int. Conf. on Knowledge-Based Intelligent Information & Engineering Systems (KES 2004).
- [12] Rinaldi F., Dowdall J., Hess M., Mollá D. & Schwitler R., 2002, Towards Answer Extraction: An Application to Technical Domains. In F. van Harmelen (ed.), ECAI 2002, Proceedings of the 15th European Conference on Artificial Intelligence, pp. 460-464. IOS Press, Amsterdam.
- [13] Roy C. B., Ravinder, K., & Roya S., 2006 The criterion-related validity of a computer-based approach for scoring concept maps, International Journal of Instructional Media
- [14] Zwaan R, & Radvansky G., 1998, Situation models in language comprehension and memory. *Psychol. Bull.* 1998 Mar;123(2):162-85. Review.