

*Trine Dahl**

Text Summarisation: From Human Activity to Computer Program. The Problem of Tacit Knowledge

Abstract

In this article I discuss whether the human activity of text summarisation can be successfully simulated in a computer. In order to write a computer program that produces high-quality summaries it becomes necessary to specify the cognitive processes involved when humans summarise text. As texts can be summarised in many different ways, evaluation of summaries becomes an important aspect in the discussion. The article discusses relevant factors in such an evaluation process. It turns out that humans when summarising texts make use of knowledge which is not readily open to scrutiny; it is tacit knowledge. This makes it very difficult to produce computer-generated summaries which are as good as those produced by skilled humans. New developments within artificial intelligence, relying on network processing techniques, may offer solutions to the problem of dealing with tacit knowledge. At present, acceptable computer summaries may be generated by programs combining accessible human knowledge of the summarisation process and knowledge about text.

1. Background

The emergence of the Internet as a commonly available information distribution channel in the 1990s intensified research efforts to develop automatic text condensation programs, as the wealth of electronically available documents made information overload an imminent danger. Automatic text condensation, like machine translation, involves the processing of natural language in a computer. In order to get the computer to perform tasks traditionally performed by human beings, programmers must be able to describe the human activity in a way the computer ‘understands’ and can imitate. Research groups working on such issues are usually found in academic departments and commercial

* *Trine Dahl*
Department of Languages
Norwegian School of Economics and Business Administration
Helleveien 30
N-5045 Bergen
Trine.Dahl@nhh.no

enterprises dedicated to the field of artificial intelligence (AI). Work involving automatic text condensation, or text summarisation as it is generally known, goes back to the early days of the computer, with Luhn's (1958) study as one of the very first attempts. Efforts to make the computer emulate the human cognitive processes involved in text summarisation soon revealed, however, that the formulation of such rules for simulating human behaviour in this very complex task is at best difficult. Today, four decades after the development of the first automatic summarisers, no automatic summarising program exists that can claim a high degree of success in condensing full texts from all domains to summary form. In the following, I will discuss whether it will ever be possible to write a computer program that matches a skilled human being's performance in text summarisation.

2. Introduction

Almost every day most of us will have to practise the skill of summarising something for other people. The full 'story' may be in the form of an event experienced, a film seen, a meeting attended, a text read, etc. Some people are good at presenting just the salient facts, while others get bogged down by irrelevant details and detours. The skill of summarising is a skill that increases with maturity and training, so that young children show poorer performance than older ones, and novice students are surpassed by mature ones (cf. e.g. Brown and Day 1983). With the development of new publishing media like the World Wide Web, access to new information has become greatly simplified compared to just a decade ago. Researchers can easily take advantage of each other's findings, as new results are made available almost immediately upon discovery. This situation has, however, created a new problem, namely how to handle all the information and how to decide its relevance. The problem of information overload has, as already stated, attracted research interest from people working within the field of artificial intelligence. Numerous research groups around the world have developed software programs that condense full text to summary form. These programs display varying degrees of success with regard to their ability to find and present the most salient information expressed in the full text. However, no current computer program can claim to match a skilled human's ability to condense text.

In order to get a computer to perform a task, very specific instructions are needed. The task must be broken down into many separate operations. Each of these operations must be described in such a concrete way as to be implementable in a computer program. One of the problems of human text summarisation is that there is no one generally accepted way of condensing a text. Experiments have shown that a group of people given a text to summarise will agree on some of the information to be included in a condensed version, but certainly not all (Salton et al. 1997). Nor are there universally accepted criteria for evaluating different summaries of a text (Gibson 1993, Jing et al. 1998, Dahl forthcoming (a)).

In this article I discuss whether the problems software engineers have with writing text summarisation programs matching human performance can be explained by reference to *tacit knowledge*. This concept is first and foremost associated with the Hungarian scientist and philosopher Michael Polanyi (1958/1962, 1967), but it has been taken up by other philosophers and given different interpretations according to their general theoretical positions. The term *tacit knowledge* has today gained currency in a number of other domains, notably knowledge management and pedagogy. Perhaps as a consequence of being used within different fields, the concept behind the term escapes easy definition. At times tacit knowledge is used about knowledge that can never be articulated, but must remain tacit. This may be called ‘inarticulate knowledge’. At other times the term represents knowledge which has become internalised through formal training and practice. Such knowledge can be verbalised and made part of conscious reasoning again. It is, so to speak, returned to the category of explicit knowledge. I also see a third phenomenon that can be covered by the term tacit knowledge. The use of the computer to emulate human behaviour has led to a sharper focus on how humans actually do things. In order to formulate rules for human behaviour, some aspects of a process that may previously only have existed as tacit knowledge can be verbalised and moved from the subconscious to the conscious level. This kind of knowledge may be called ‘unarticulated knowledge’, alluding to the potential for verbalisation.

The fact that some kinds of tacit knowledge must remain tacit has implications for automatic text summarisation. It is very likely that future computer programs may yield higher quality output than today’s

programs, by for instance verbalising some previously tacit knowledge (habits and unarticulated knowledge), but since some knowledge cannot be conscious (since it is inarticulate knowledge), it seems unlikely that a computer will ever be able to match high-quality human performance in text summarisation.

3. The concept of Knowledge

The invention of the computer in the 1940s led to a revival of the view within philosophy of human beings as machines. Such a view implies that human behaviour is seen as rule-based and may be transferred to a computer which is then able to mimic this behaviour. This view is known as the computational view of the mind. According to such a view, what counts as progress within artificial intelligence boils down to putting more – rule-based – knowledge into the machine. Early success in computer programming gave nourishment to this line of reasoning. As machine power increased in the 1970s, expert systems were developed that seemed to work just as well as, or sometimes even better than, human experts (Feigenbaum 1984). Such programs are dedicated to solving problems and giving advice within a specialised area of knowledge, a micro-world. The systems are developed by eliciting expert knowledge from a group of outstanding practitioners in the relevant domain. This knowledge is then formulated as rules. One program that can claim to be successful in the sense that its decisions were judged to equal that of a human expert is MYCIN. The program, which was developed at Stanford University, is an interactive system that assists physicians in diagnosing infectious diseases such as meningitis. Another successful program is LOKE (Sørensen and Nordhus 1988). This system was developed by engineers and knowledge experts at Statoil, the Norwegian state-owned oil company. LOKE was constructed to assist the drilling supervisor in the drill bit selection process.

However, other areas of human behaviour turned out to be far less amenable to computerisation. Efforts to formulate rules for tasks involving natural language processing yielded very disappointing results. The proponents of the computational model of the mind claim that setbacks in these fields are temporary only, and that once we are able to formulate the knowledge that is ‘hidden’ from inspection today, there is no reason to believe that computers will not be able to match the performance of humans in natural language processing in the future.

Opponents of the idea of equating humans with machines, such as the American professor of philosophy Hubert Dreyfus (1979, 1992) and John Searle (1980), maintain that the problem is a much more fundamental one. To represent all of human knowledge as a mere symbol manipulation system gives an insufficient account of human understanding. Support for this position can also be found in the account given by one of the developers of the LOKE expert system. In the knowledge elicitation phase, the programmers became frustrated when the domain experts were unable to give precise answers to questions about their own selection criteria. One of the participants in the project states the following:

Som i andre prosjekt fikk vi en del problem under veis (sic). Et hovedproblem var at mange av de svarene vi fikk ikke var noen skikkelige svar. **De inneholdt liksom ikke kunnskap i vår forstand av ordet.** Man måtte vel kunne si noe mer fornuftig om hvorfor boringen går greitt enn at man merker det på ristingen i riggen? (Sørensen n.d., emphasis added)

What is knowledge, then? When we think about knowledge we tend to focus on what goes on at the conscious level. Especially in a situation of formal training, we emphasise the transfer and reproduction of facts and beliefs. This is particularly pertinent to academic educations, but also to some extent to apprenticeships in various crafts. The apprentice, in addition, clearly needs to pick up skills that will be more difficult to pin down as actual knowledge. The same, in fact, applies to academic settings. Knowledge about genre and register in academic writing may be subconscious, as may our way of arguing a point in a piece of research. To me it seems obvious that knowledge is not only something that can be acquired through a book or a lecture. We are socialised into research communities through training at the conscious as well as the subconscious level. The British sociologist H.M. Collins puts it this way: ‘The general rule is that we know more than we can say, and that we come to know more than we can say because we learn by being socialized, not by being instructed.’ (Collins 1990: 7-8).

This idea of ‘knowing more than we can say’ can be linked to the view of knowledge developed by Polanyi (1958/1962, 1967). Polanyi sees a tacit dimension to knowledge. The knowledge that is open to scrutiny – often referred to by terms like *propositional* knowledge or *explicit* knowledge – rests upon this tacit knowledge. If we support this

view that knowledge is something more than explicit facts and beliefs that can be formulated verbally, this will have implications for our view of artificial intelligence. How can a machine ever come close to intelligent behaviour if much of human knowledge turns out to be impossible to verbalise? Traditional artificial intelligence literature tends to disagree with these ideas (for another view within AI, see the discussion of *connectionism* in section 4). In programming involving natural language processing (notably machine translation and text summarisation), the limitations of most systems are usually ascribed to deficiencies in the knowledge base that the program draws on, or linguistic problems like word ambiguity and anaphora resolution. Success programs involving manipulation of natural language tend to apply to very restricted areas. A case in point is the Canadian TAUM METEO program that translates weather forecasts between English and French. Another program, which is still being refined, is a communications system for use in the Channel Tunnel. The program, called PoliceSpeak, is a bilingual text messaging system. It allows French and British Police to exchange operational text messages in their own languages. In a comment on the eventual success of the system, one of the developers, Edward Johnson, says the following:

The success of strategies like PoliceSpeak depends ultimately on the abilities of the service professionals themselves and considerable mental discipline is required of them. They need to tolerate the tension between their explicit operational language and their everyday language in which communicative power seems to lie not so much in exactness but in fuzziness of expression and tolerance of ambiguity; in which very little has universal hard and fast 'currency' and through which feelings and information are exchanged by way of a mysterious process of linguistic negotiation during which things are said which are not meant; things are successfully conveyed which have not been adequately articulated and things agreed which have never been stated. (Johnson 1994: 745)

The last part of the quote again points to the problem raised in this article: Is it possible to capture all aspects of human behaviour and formalise them for use in a computer?

Knowledge presupposes intelligence. In the next section I will discuss this concept in relation to human beings as well as computers.

4. Intelligence in humans and machines

An attempt to define intelligence can never be more than an attempt. We cannot hope to grasp all the aspects of such a complex notion. A definition of intelligence should include the capacity to learn. In order to learn, human beings exploit their ability to see, use language and motor facilities. These facilities combine in the use of knowledge, both common-sense and domain-specific, to acquire new knowledge. Another very basic aspect of human intelligence is creativity. Both the capacity to learn and the aspect of creativity must be included in our definition of what it means to be a human being. Until people used both their capacity to learn and their creativity to invent the computer in the early 1940s, the study of intelligence was a field dominated by philosophers and psychologists. The computer changed this and spawned a brand new field of research activity, termed *artificial intelligence* (AI). The phrase was coined by John McCarthy at the famous Dartmouth Conference in New Hampshire in 1956, which is considered as the birth of the discipline. The research field today encompasses disciplines such as computer science, cognitive science, linguistics, psychology and philosophy.

To many people the term *artificial intelligence* is an oxymoron. They maintain that only humans can display intelligence. The term is nevertheless defined as the science of designing machines and systems which do things that would require intelligence if done by people. Within the domain of artificial intelligence we distinguish between the strong AI and the weak AI programme. Proponents of strong AI think that we will eventually be able to produce real intelligence in a machine. This implies that we can create a machine that will be a thinking entity, a conscious being. Those who support the weak AI programme think that the creation of truly intelligent machines is an unrealistic goal, and go for the more modest aim of producing very good emulations of intelligent behaviour.

In order to get the computer to emulate human behaviour it becomes necessary to find out about the workings of the brain. The human brain has been studied from many different angles and today we know much about the way neurons and neurotransmitters operate. But when we try to copy the workings of this so-called 'wetware' and transfer it to computer hardware, it becomes abundantly clear that much about the brain remains a mystery. One important finding in this respect is that a piece

of knowledge does not seem to be related to one specific location in the brain; rather, a combination of different locations is involved. The philosopher Jack Copeland, in a comparison between human and machine memory, states that '[e]ach piece of information in a computer's memory has a *single* address or 'handle'. In contrast, the human brain is capable of using any one of an open-ended collection of descriptions to retrieve one and the same fact.' (Copeland 1993: 188).

As already mentioned, the mimicking of human activity has been successful in some areas, while in other fields the researchers are left disillusioned about the computer's abilities, or rather people's ability to formulate the necessary requirements for a program to perform tasks to the same high quality as human beings. One of the most active proponents of the strong AI programme, Douglas B. Lenat, states the dilemma of AI in the following way: 'One of the most frustrating lessons computers have taught us time and time again is that many of the actions we think of as difficult are easy to automate – and vice versa.' (Lenat 1995). To Hubert Dreyfus (1992) this is proof that traditional AI is a degenerating research programme. We have a situation with a scientific enterprise starting out with great success in a limited domain. Attempts to apply the approach more broadly, however, are less successful. Important phenomena cannot be handled by the techniques stipulated, and the research programme will eventually be abandoned. This, according to Dreyfus, sums up the current situation within traditional artificial intelligence.

However, recent developments within the branch of artificial intelligence research known as *connectionism* have generated new optimism with regard to the potential for emulating human behaviour. While traditional AI research focuses on how a single computer can be programmed to perform symbol processing, the connectionist approach involves the use of a network of processing devices, meant to imitate the workings of the human brain. This approach operates under various names, such as *connectionism* (cf. e.g. Churchland and Sejnowski 1992), *parallel distributed processing* (PDP) (cf. e.g. Rumelhart and McClelland 1986) and *artificial neural networks* (relevant research information can be found at <http://www.ewh.ieee.org/tc/nnc/research>). The terms are associated with different research environments and display some theoretical differences, but the many common features justify a common treatment in the present context. As the term *connec-*

tionist seems to be the most general term, it will be used in the following. The connectionist approach relies on the interconnection of several simple processing units in order to produce the kind of complex behaviour associated with the human brain. Thus, rather than approaching concepts and actions in terms of mere symbol processing expressed in algorithmic form, connectionist approaches are inspired by the workings of the brain. This implies that in a connectionist system there are connections between several processing units. These connections can be modified (strengthened or weakened) in order to accommodate new information. In most cases such systems include a learning procedure. The system can be trained on empirical data of the behaviour we are trying to emulate. In the training process, structures from the existing data can be picked out and exploited to modify the connections between the processing elements. The system is hence capable of learning patterns from examples. Much research effort is currently invested in projects taking a connectionist approach to computing. Such systems have seen encouraging progress in areas involving e.g. sensory data, including speech recognition. Another area involving language is signature analysis, where handwriting recognition can be obtained through connectionist systems. Other areas of application include the health care industry, the financial industry and manufacturing.

To my knowledge, little if any connectionist-inspired research to date has involved the natural language processing activity of text summarisation. Hence, the discussion of this task in the next section focuses on results obtained in projects taking more traditional AI approaches to text summarisation.

5. Text summarisation

A very simple definition of what the product of a text summarisation process is could be put as follows: a shortened version of a source (original) text, containing the main content of that text. Several terms are in use for such a derived text, such as *summary*, *abstract* and *extract*. *Summary* and *abstract* are often used as synonyms for a derived text based on a source text, while an *extract* is a derived text consisting of fragments of the source text (cf. Dahl forthcoming (b) for a more thorough discussion of the three terms). In human text summarisation two different situations occur depending on who is doing the summarising.

One common situation is the one where the summariser is not the author of the source text. This is the case in pedagogical summarising activity and in the abstracting and indexing services. The cognitive process first called into action in such cases is source text comprehension. This aspect is of course particularly relevant with texts from special knowledge domains. Another important – and interdependent – skill is the summarising skill itself, involving the picking out of relevant content for inclusion in the summary as well as the condensation aspect, turning local information into globally relevant material. Finally, linguistic ability is needed to present the information in a readable form. Another situation presents itself when the author of the source text is summarising his or her own text. In that case the comprehension phase is irrelevant, but the other two still apply.

In automating this process the computer must be likened to a secondary abstractor (that is, someone other than the author of the source text), bringing the comprehension phase to the fore. How can we make the machine ‘understand’ the source text? Obviously an extremely comprehensive knowledge base is required if the summariser is meant for non-restricted use. This brings us to the common-sense knowledge problem in artificial intelligence. Even young children know a lot of things about the world which the digital computer is ignorant of until the knowledge has been explicitly stated. If a child is told that mummy is in Copenhagen, the child also knows that mummy’s arms are in Copenhagen. A computer, when asked about this, is not able to infer the correct answer unless it has been specified in the program. In order to address this very serious problem of lack of common-sense knowledge, Douglas B. Lenat in 1984 initiated a large-scale project, called CYC (<http://www.cyc.com/>), aiming to develop a large computational database and search tools which will enable artificial intelligence systems to access the knowledge making up common sense (Lenat 1995). The approach used to gather this common sense information is to capture all the knowledge – both implicit and explicit—in a hundred randomly selected articles in the Encyclopaedia Britannica. The CYC project can be regarded as a start in the right direction when it comes to common sense knowledge. In addition, large amounts of domain-specific knowledge will be needed for many texts. The nontrivial nature of this task is obvious.

The comprehension problems caused by the lack of common-sense as well as domain-specific knowledge represent the first major problem in automatic text summarisation. Taking such a deep approach to text condensation implies not only generating a full semantic representation of the source text. In a next, perhaps equally difficult, step explicit rules for how the essential information can be selected from the source text must be formulated. Then, a semantic representation of the new, condensed, version, must be generated before this version is given a final, textual, form (cf. e.g. Sparck Jones 1997). To date no automatic summariser drawing on a knowledge base has been developed that can handle unrestricted text. A few examples exist for very limited domains like, for instance, the summarisation of banking telexes (Young and Hayes 1985).

Any problems the summarising program has in the comprehension phase will be carried over into the actual summarisation phase. How can the program find the essential information when global understanding of the source is missing? In order break this deadlock, some AI researchers have teamed up with text linguists to skim as much as possible from the surface of the source text. Textual factors that can be exploited in such an approach are, for instance, overt rhetorical structure indicators like section headings in research articles and the lead in newspaper articles, or indicator phrases in articles such as 'The principal aim of this paper is to investigate...', 'This paper reports the results of an experiment...', and 'It has been shown that...' (Paice 1981). Another approach has been to use significant lexical cohesion between sentences to trace the essential information in a text. The lexical cohesion method has yielded quite promising results (Hoey 1991, Benbrahim 1996, Barzilay /Elhadad 1997). Current research projects on automatic text summarisation (e.g. Teufel /Moens 1998, Dahl forthcoming (b)) suggest that a combination of such methods might be a way forward towards better summaries.

A common characteristic of such shallow (as opposed to the deep) automatic summarisers using textual information is that the output text is an extract rather than an abstract. What the programs do is lift sentences from the source text and present them in concatenated form as a summary. Hence, in contrast to abstract writing, no synthesis of information takes place. A problem with this approach is the potential lack of coherence in the summary since local coherence may be lost in the

extraction process (this is usually clearly demonstrated in the use of Microsoft's Word 98 AutoSummarize facility).

The followers of the strong AI programme reject this idea of surface text-based summarisers, and keep their focus on developing more powerful knowledge bases and the formulation of rules to account for summarisation behaviour. However, they recognise that the development of automatic summarisers based on deep approaches to the task is a long-term goal (Sparck Jones 1999).

6. Evaluation of summaries

As indicated in the introduction, the field of automatic text summarisation is an active one, and particularly the surface-text based programs may claim some success. This brings us to the thorny question of evaluation. A common conclusion to many research projects in text condensation is that there is surprisingly little agreement among a group of human summarisers of how a full text can be condensed (Salton et al. 1997). The same applies to the assessment of summaries by a group of people (Gibson 1993). These two situations both reflect the most essential aspects of text summarisation: The first concerns the process of condensing a text; the second concerns the product, the summary itself. The process aspect is reflected in the following question: If we are to mimic human behaviour in the automatic summarisation process, what kind of human behaviour should we use as our model? As regards the product aspect, the following question must be asked: How can we measure the success of each program's output and say that one program works better than another? In the following discussion of relevant criteria in the evaluation of summaries, the process and product aspects are both important. It is not obvious how they can be kept separate. A description of mimic-worthy summarisation behaviour seems almost inextricably linked to our evaluation of the summary produced. In both instances some of the factors involved remain impossible to express in words; they operate at the tacit level of knowledge.

We might start by setting up two broad categories of evaluation criteria: they may be *quantitative* or *qualitative* criteria. A purely quantitative criterion is the time it takes to produce a summary from a source text. Today production time for automatic summarisers ranges from several hours to a few seconds. This is obviously important in a com-

mercial context. The length of a summary (can it be easily customised to the requirements of the user of the summary?) is another quantitative criterion which can be imposed in an evaluation procedure.

As for qualitative criteria, one method of evaluation could be to formulate some kind of norm that the summary could be measured against. For certain types of text there are such norms in the form of officially recognised standards. A case in point is the ISO 214-1976 (E) standard for writing abstracts. Abstracts of journal articles within the natural sciences tend to follow this standard quite closely, while abstracts of research articles within the humanities are much less standardised (Tibbo 1992). If we accept a standard like the ISO 214 as a valid and relevant one – and this of course is the point about having agreed standards – evaluation of a summary of a research article can be done by measuring how the derived text matches the recommendations stated in the standard. At some high level of evaluation this is useful. The ISO standard defines an abstract as ‘an abbreviated, accurate representation of the contents of a document, without added interpretation or criticism’. It recommends that the abstract ‘states the purpose, methodology, results, and conclusions presented in the original document’. Using this norm, it might be possible to point to a missing concept/idea in a summary, or indicate criticism included, and evaluate it as flawed. Measuring against such a general norm can, however, only take us part of the way. There will be many cases where the norm has been followed, but still human evaluation of different summaries of a source text will vary. This is demonstrated in studies by for instance Gibson (1993) and Salton et al. (1997). Gibson (1993) states that:

it seems reasonable to conclude that the success of an abstract is determined not by any one single factor, but by a complex set of factors. Further, it seems reasonable to conclude that informants [=evaluators. My note] either do not agree on the composition of these factors, or that they choose to attach differing levels of importance to each. (Gibson 1993:122)

Human evaluation is regarded as the most powerful qualitative evaluation criterion today, but as we have seen, also the most elusive.

The measures available for evaluation that I have touched upon so far, that is production time, length, benchmarking against a standard and human preference, were all presented without any reference to contextual factors. The most influential contextual factors are *audience* and

use. A human summariser must take readership into account when producing a derived version of a source text. This is particularly relevant in the communication of specialist knowledge. In symmetrical communication situations, where the readership of the source text and the summary share the same knowledge base, no special adaptation strategies are needed in the summary production phase. In asymmetrical situations, like for instance popularised accounts, much effort must be put into converting the source text content into a suitable form in the summary. A special adaptation phase is introduced in the summarisation process. Adaptation can take place at various levels. Explicitation of content to replace inferences may be a relevant strategy, as may replacement of special terminology with general language words.

The second contextual factor – use – requires the summariser to pay attention to what the summary is meant for. Relevant categories of summary use are: as a *replacement* of the source text, as a *relevance indicator* for the source text or as a *content pre-organiser* of the source text.

So, is it relevant to apply such evaluation criteria to an automatic summariser, and can we expect it to handle any of the contextual factors? The quantitative criteria, such as production time and length, have been shown to be relevant. As for the qualitative criteria, it should be feasible to write a program that adheres to the standard for writing abstracts by including information from all the parts of e.g. a research article. The problem of how to pick the most relevant information from each part remains, however.

As for adaptation to audience and use, to my knowledge none of the existing computer programs may claim to accommodate such parameters. What we are left with is that indefinable notion of human individual judgement. This is not unique to the production of summaries. It is a general issue related to text production and evaluation, as for instance Hertzberg (1995) has shown. We meet it in the production of all kinds of texts from novels to technical reports. We also meet it in derived texts like translations, text adaptations and summaries. The fact that text production in general and text summarisation in particular is so difficult to explain and evaluate implies that the cognitive processes involved are so complex that it does not make sense to try to explain them by reference to linguistic phenomena alone.

There have been attempts at finding out what mental processes are involved in human text summarisation, with the aim to implement those aspects in a computer program. One such project taking a psychological approach is that of Endres-Niggemeyer et al. (1995). The method they employ in discovering the processes involved is thinking-aloud protocol. The human summariser tape-records what he or she is doing during the summarisation process and why. The intention is to capture the routine working processes involved in this task. Endres-Niggemeyer et al. (1995) remark that :

[n]atural working steps demonstrate intact and competent human behaviour. **An important part of human know-how lies in the skilled combination of different strategies.** Only in working steps can we observe how successful abstracting acts are performed and how several pieces of know-how are integrated to achieve a cognitive aim. (Endres-Niggemeyer et al. 1995: 636, emphasis added)

As is the case with many such projects, no computer implementation has taken place yet. In addition, only four working steps are worked out, aimed at producing a topic sentence for the summary. There is quite a way to go from this rather modest start to the production of a complete summary.

7. The indefinable: tacit knowledge

The definition of artificial intelligence implies that its research domain is the emulation of human behaviour. When it comes to tasks like text summarisation, the lack of adequate common-sense knowledge bases has led to research effort being channelled into text extraction, drawing on textual knowledge, rather than the pursuit of ‘real’ text summarisation with synthesis of content and coherent summary output. The attempts that have been made at writing such text-based programs have managed to capture some of the features that humans probably make use of. Research on writer – reader interaction (e.g. Hoey 1991, Myers 1992) shows that we are alerted to important points in a text through the use of indicator phrases and repetition of terminology and phraseology, that is, lexical cohesion. The existence of standards like the ISO 214 has also had a strong normative effect, at least within the natural sciences.

But even if very powerful common-sense and domain-specific knowledge bases were to be developed, this would still not be sufficient. The researchers within the AI domain – particularly the adher-

ents of the strong programme – see human beings as machines. Any difficulty in formulating the rules for required output is temporary only. If we disagree with this view of humans and support the ideas of Dreyfus (1972, 1992) and Searle (1980) that we are much more complex entities, we cannot hope to write computer programs that attain the level of performance of skilled humans in all tasks. Natural language processing may well turn out to be such an ‘impossible’ task.

The consequence of regarding humans and machines as different sorts of entities is not necessarily that we should abandon the computer in all natural language matters. On the contrary, by continuing the effort of writing better programs to manipulate natural language, we learn much about the human activity involved. This way human performance improves at the same time through conscious focusing on the processes involved. Some of the aspects involved in natural language processing clearly belong to the area of tacit knowledge. We recognise this through the quotes from Johnson (1994), Gibson (1993) and also, in fact, Endres-Niggemeyer et al. (1995). These quotes effectively imply that how we produce a text and how we evaluate it depend on factors which are not readily open to scrutiny. Some of these factors may in future be verbalised and hence transferred to the category of explicit knowledge. Such knowledge can be assigned to the category of currently unarticulated knowledge. In the conclusion to their paper, Endres-Niggemeyer et al. (1995) allude to such a situation. The empirical model used in their study is said to uncover ‘features of the intellectual process which **had not been specified earlier**’ (1995: 671-672, emphasis added). We will, however, be left with aspects which can only be described by phrases like ‘Text 1 *reads better* than text 2’, ‘This version is *more lucid* than that one’, ‘This is a better summary of text A than that one’, etc. What is actually implied in such evaluative expressions remains hidden in the interface between the human being uttering them and the socio-cultural context this person exists in. Such knowledge remains tacit; it is inarticulate knowledge. As for the branch of artificial intelligence research involving connectionist approaches to the computation of human activities, this may be a way forward. The Canadian philosopher Paul Thagard, for one, points to connectionism as a potential way to handle even the tacit – inarticulate – component of knowledge. Speaking against what he sees as the excessive optimism of proponents of the

symbol-processing approach, he expresses his own views on the future of AI in the following quote:

My own guess is that we are still missing a number of key conceptual ingredients for understanding natural and artificial intelligence, so I would project centuries of research rather than decades. It is also possible that the problem of understanding the mind is too complex for the mind to solve, or that the computational approach is fundamentally flawed, perhaps because thought is ineluctably tacit or situated in the world. But AI has at least the possibility of modeling tacit knowledge (through connectionist distributed representations or cases used analogically) and of situating cognition in the world (through distributed AI and robotics). Progress has been sufficient in the past few decades that the only reasonable collective scientific strategy is to wait and see how the ideas play themselves out. (Thagard 1993)

Predictions about the future are notoriously hard to confute. In my opinion, however, humans and computers are different entities and will remain so, even as new and more sophisticated machines and networks are developed.

References

- Barzilay, Regina & Elhadad, Michael (1997). Using lexical chains for text summarization. In Mani, I. and M.T. Maybury (eds.) *Proceedings of the Workshop on Intelligent Scalable Text Summarization*, in association with ACL/EACL'97.
- Benbrahim, Mohamed (1996). *Automatic text summarisation through lexical cohesion analysis*. Unpublished PhD thesis, University of Surrey.
- Brown, Ann L. & Day, Jeanne D. (1983). Macrorules for summarizing texts. In *Journal of Verbal Learning and Verbal Behavior* 22. 1-14.
- Churchland, Patricia S. & Sejnowski, Terrence J. (1992). *The computational brain*. Cambridge, Mass.: MIT Press.
- Collins, H.M. (1990). *Artificial experts. Social knowledge and intelligent machines*. Cambridge, Mass.: MIT Press.
- Copeland, Jack (1993). *Artificial intelligence: a philosophical introduction*. Oxford: Blackwell.
- Dahl, Trine (forthcoming a). *Text evaluation methods: The case of computer-generated summaries*.
- Dahl, Trine (forthcoming b). *Lexical cohesion-based text condensation: An evaluation of automatically produced summaries of scientific articles by comparison with author-written abstracts*. PhD thesis.
- Dreyfus, Hubert L. (1979). *What computers can't do: the limits of artificial intelligence*. New York: Harper & Row.

- Dreyfus, Hubert L. (1992). *What computers still can't do: a critique of artificial reason*. Cambridge, Mass.: MIT Press.
- Endres-Niggemeyer, Brigitte, Maier, Elisabeth, & Sigel, Alexander (1995). How to implement a naturalistic model of abstracting: Four core working steps of an expert abstractor. In *Information Processing & Management* 31 (5). 631-674.
- Feigenbaum, Edward A. (1984). *The fifth generation: artificial intelligence and Japan's computer challenge to the world*. London: Michael Joseph.
- Gibson, T.R. (1993). Towards a discourse theory of abstracts and abstracting. Monographs in systemic linguistics, 5. University of Nottingham.
- Hertzberg, Frøydis (1995). Uttalte og uuttalte normer for vitenskapelig skrivning. In Johnsen.E.B. (ed.) (1995). *Virkelighetens forvaltere*. Oslo: Universitetsforlaget. 187-205.
- Hoey, Michael (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
- ISO 214-1976 (E). (1976). International Organization for Standardization – Documentation: Abstracts for publications and documentation. Geneva.
- Jing, Hongyan, Barzilay, Regina, McKeown, Kathleen & Elhadad, Michael (1998). Summarization evaluation methods: Experiments and analysis. In Hovy, E. and D. Radev (eds.) (1998) *Intelligent Text Summarization*. Papers from the 1998 AAAI Symposium, Stanford, California. Technical Report SS-98-06. Menlo Park, Ca.: AAAI Press. 51-59.
- Johnson, Edward (1994). PoliceSpeak. In Brekke et al. (eds.) (1994). *Applications and implications of current LSP research*. Bergen: Fagbokforlaget. 738-745.
- Lenat, Douglas B. (1995). Artificial intelligence. A crucial storehouse of commonsense knowledge is now taking shape... In *Scientific American*. September. 80-82.
- Luhn, Hans P. (1958). The automatic creation of literature abstracts. In *IBM Journal of Research and Development* 2 (2). 159-165.
- Myers, Greg (1992). 'In this paper we report...': Speech acts and scientific facts. In *Journal of Pragmatics* 17. 295-313.
- Paice, Chris D. (1981). The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. In Oddy et al. (eds.) (1981). *Information retrieval research*. London: Butterworths. 172-191.
- Polanyi, Michael (1958/1962). *Personal knowledge. Towards a post-critical philosophy*. Chicago: University of Chicago Press.
- Polanyi, Michael (1967). *The tacit dimension*. London: Routledge & Kegan Paul.
- Rumelhart, David E., McClelland, James L. & the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (2 vols.). Cambridge, Mass.: MIT Press.
- Salton, Gerard, Singhal, Amit, Mitra, Mandar, & Buckley, Chris (1997). Automatic text structuring and summarization. In *Information Processing & Management* 33 (2). 193-207.

- Searle, John (1980). Minds, brains and programs. In *Behavioural and Brain Sciences* 3. 417-424.
- Sparck Jones, Karen (1997). Summarising: Where are we now? Where should we go? Opening talk, Madrid workshop on intelligent scalable text summarization.
- Sparck Jones, Karen (1999). Automatic summarising: factors and directions. In Mani, I. and M.T. Maybury (eds) *Advances in automatic text summarisation*. Cambridge Mass.: MIT Press.
- Sørensen, Terje (n.d.). Kunnskap og teknologi. En vandring i et kunnskapsmessig landskap. <http://www.pakt.ntnu.no/yearbook/bok95/artikler/kunnskap.htm> (site visited 14 January 1999).
- Sørensen, T. & Nordhus, O. (1988). Loke: a drill bit selection system. In Moralee (ed.) (1988). *Research and development in expert systems IV*. Cambridge: Cambridge University Press. 32-41.
- Thagard, Paul (1993). Societies of minds: Science as distributed computing. <http://cogsci.uwaterloo.ca/Articles/Pages/Societies.html> (site visited 10 January 1999).
- Teufel, Simone & Moens, Marc (1998). Sentence extraction and rhetorical classification for flexible abstracts. In Hovy and Radev (eds.) (1998). *Intelligent text summarization*. Papers from the 1998 AAAI Symposium. Menlo Park, Ca.: AAAI Press. 16-25.
- Tibbo, Helen R. (1992). Abstracting across the disciplines: a content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online information retrieval. In *Library and Information Science Research* 14 (1). 31-56.
- Young, S.R. & Hayes, P.J. (1985). Automatic classification and summarization of banking telexes. In *Proceedings of the 2nd Conference on Artificial Intelligence Applications*. New York, NY: Institute of Electrical and Electronics Engineers. 402-408.

Web Sites

- The Cycorp Home page*. <http://www.cyc.com/> (site visited 9 December 1999).
- The IEEE Neural Network Council Home Page* (research). <http://www.ewh.ieee.org/tc/nnc/research> (site visited 13 December 1999).

WHEN TALK IS A SCIENCE...



Linguistics and Language Behavior Abstracts

*Comprehensive, cost-effective, timely coverage of current
ideas in linguistics and language research*

Abstracts of articles, books, and conference papers from nearly 1,500
journals published in 35 countries; citations of relevant dissertations as
well as books and other media.

Available in print or electronically through the Internet Database Service
from Cambridge Scientific Abstracts (www.csa.com).

Contact salar@csa.com for trial Internet access or a sample issue.

Linguistics & Language Behavior Abstracts

Published by Cambridge Scientific Abstracts

7200 Wisconsin Avenue • Bethesda MD 20814 • USA

Tel: +1 301-961-6700 • Fax: +1 301-961-6720 • E-mail: salar@csa.com
