Summary and review of the paper *Generating Summaries from Event Data*

Fredrik Olsson

October 2, 2003

1 Introduction

This paper is written as a part of the first, mandatory assignment on the course in information access, taught at the Swedish National Graduate School of Language Technology during the fall semester of 2003. The present paper is concerned with the first part of that assignemnt, to write a summary of a selected paper, in this case it is *Generating Summaries from Event Data* by Mark T. Maybury (1999).

2 Summary of (Maybury, 1999)

This section contains a summary of Maybury's paper, and the sections below reflect the structure of his paper.

2.1 Introduction

Maybury starts out by describing an effective summary as one that; *... distills the most important information from a source (or sources) to produce an abriged version of the original information for a particular user(s) and task(s).* This sentence pretty much contains the keys to the rest of the article since it have bearings at the four main types of processing involved in producing a summary: (1) analysis of source text, (2) identification of important knowledge in the source, (3) condensation of information, and (4) (surface)generation of the summary.

2.1.1 Source interpretation

Maybury points out that simple techniques will get us far, but not far enough. The *source interpretation* section in his paper contains references to previous work in the area, approaches utilizing techniques and methods such as: selecting a number of words from the first sentence(s) of a text; using clue phrases, scripts, frames or rules to detect important material in the source; combining techniques such as, e.g., word frequencies and syntactic information. On the knowledge-intense side of the spectrum, Maybury points to work carried out by Karen Sparck-Jones (1993) where she performed several types of analyses by hand (using Rhetorical Structure Theory, story grammars, predicate-argument

structures, etc), the outcome of which suggests that summarization (or generation) would benefit from a mixture of techniques, rather than the employment of any technique in isolation.

Once a source has been analyzed and transformed into some manageable structure, the parts of the structure that carry important information needs to be identified. This can be accomplished, more or less accurately, with several methods, e.g.; word-frequency distributions, clue phrases, or by reasoning about the rhetorical structure of the source text. Determining what is important is influenced by the source as well as by the information need and task at hand, the properties of the user, the discourse history, and the desired product or outcome.

The selected information may need to be represented or presented more compactly, that is, to be condensed. This can be achieved either by abstraction of aggregation.

2.1.2 Summary generation

The final step in summarization is to present the summary to the user. Few systems use text generation for rendering and presenting summaries, it is more common to use extraction and present pieces (possibly re-arranged) from the source text to the user.

Selection, condensation and generation may be done from intermediate representations, with or without references to the source text. Maybury describes experiments with two systems, one of which has access to the source text from which the events to summarize were extracted. His approach utilize both domain dependent and domain independent strategies for information selection and aggregation.

2.2 Techniques for event summarization

The focus of Maybury's paper is on how to summarize events. Information about an event may be extracted from text, entered maually, or originate from real-world (or simulated) sensors. He then turns to present an (incomplete) taxonomy of techniques available for summarizing events:

Select

1. Importance (of events). A domain specific measure of the relative significance of an event, which can refer either to individual events, classes of events, or to the characteristics thereof. The **strengths** of the technique is that it's claimed to be easy to implement. The **weaknesses** include domain and context dependence.

   Importance is not only domain or context dependent, it is also in the eye of the beholder, considering the user's needs and perspective on the event(s) to be summarized.

2. Importance (of event attributes). See above!

3. Link analysis. Domain independent technique considering frequency analysis of (semantic) relations between events. Assumption is that, e.g., events that cause or enable many other events are more significant than

events that occur in isolation. **Strengths** include domain independence, while **weaknesses** include assumption that isloated events are unimportant.

4. Statistical analysis. Assumption: events that occur frequently tend to be less significant than those that occur infrequently, while, in the analysis of the source text, events that are mentioned more frequently are more indicative of the source content and therefore more significant. Frequency alone is not enough, events' distribution over time also important. **Strengths** include easy to implement, computationally inexpensive, while **weaknesses** include requirement of large number of events to be able to produce statistically significant results.

Condense

1. Abstraction. Take a series of events and replace them with a single event. Can be accomplished by, e.g., pattern matching or plan recognition. Leaving out information reduces detail while, at the same time, potentially increases ambiguity. **Strengths** include enabling summarization across events, while **weaknesses** include complex reasoning about abstraction, losing precision and coherency.

2. Aggregation. Similar events can be unified into a single description. **Strengths** include small presentation space, while **weaknesses** include loss of details and the requirement to define and identify similarities between events.

Present

Presentational techniques can be used for shortening the length of time and space required to express information. Examples of techniques are: exploitation of previous discourse history; selection of other media than pure text to present the information to the user, such as graphics and tables; and exploitation of rhetorical or intentional structure to select what to include or exclude in the summary. **Strengths** include shortened time to convey information to the user, while **weaknesses** include reasoning about context, media and rhetoric.

2.3 SumGen: A summary generator

The primary setting in which Maybury's system work is that of an object-oriented battle simulation system (he also presents results on applying the system to the joint-venture scenario used in MUC-5, see Section 2.4). The task of the summarization system, SumGen, is to take as input time stamped event messages from the simulator from which it selects, aggregates, and presents a battle summary. A run of the simulator may result in several thousand event messages.

First, event messages are parsed into a semantic network of events, in which each event contains attribute slots for, e.g., type of event, its time and place, and role slots indicating, e.g., the agent and recipient of the event.

Selection of events from the network is based on an importance metric measuring the relative significance of a given event. It is a function of three things:

1. the frequency of occurrence of events in the network,

2. the type and amount of relations associated with an event in the network, and

3. domain specific knowledge of what is important.

Once SumGen has determined what events are to be in the summary, the events are condensed using aggregation; equivalent events as well as similar events are aggregated. The original events are then replaced by the aggregated ones, grouped by the mission they are related to, e.g., air strike or refueling.

The overall organization of the summary is accomplished using a communicative-act based text generator that sequences the report first by topic, and then chronologically within topic. The resulting report may still be to lengthy for some users, but it is possible to configure the SumGen system in such a way that it produces summaries targeted at stereotypical users interested in more specific characteristics of particular events (e.g., a logistician is mainly interested in the logistics carried out in a simluated battle, and not in missile launches).

2.4 Evaluation

The SumGen, as applied both to the battle simulator and to the joint venture scenario of MUC-5, was evaluated in an experimental study involving 22 subjects. The subjects were to read long and summarized battle simluation reports and to retrieve the names, participants, time and duration of all missions that appeared in the text.

As for the joint venture domain, the subjects were asked to extract information from 11 randomly selected texts (out of 748 Wall Street Journal documents) considering both the original information source as well as the summarized versions of those. The information to extract was the type, partners and status of all joint ventures mentioned in the texts.

Maybury points out that the evaluation of the battle simulator was more important than the evaluation of the joint venture scenario, simply because the former also involved both selection and condensation, whereas the latter assumend that those tasks had already been solved by an information extraction component.

I will not give any number of the evaulation in this summary of the paper, since it would basically mean that I would need to re-write the whole evaluation section from Maybury (1999) to get all figures right. Instead, I will settle for reproducing the general tendencies reported by Maybury. The evaluation shows that

- the average time for solving the tasks was lower on the summarized documents,

- the average document length was (obviously?) shorter for the summaries as compared to the original documents,

- the average sentence length was shorter in the summaries,

- the average precision and recall was higher when subjects solved the tasks using the summaries rather than the full text documents.

The techniques used have not been validated in more than the reported two domains, but they are claimed to be domain independent and should thus be feasible to use in many other event-oriented applications.

2.5 Future research

The contents of the future research-section essentially boils down to the following four issues:

1. Evaluation remains important research area.
2. What makes a good summary for who and for what purpose remains an important research issue.
3. Use of more linguistic techniques remains an important research issue.
4. Combination of techniques appears to be good for analysis, and the same might be expected for generation.

3 Review of (Maybury, 1999)

It is clear that the issues brought up in (Maybury, 1999) apply to more than just the issue of producing summarizes from events. Over all, the paper is well worth reading, although it suffer from a touch of wordiness.

3.1 Pros

User's are different. I like the way Maybury, on several occasions, advocates the view that the relvance of a summary is very much in the eye of the beholder.

Mixed-technique approach. From where I come from, it seems obvious that applications such as summarizers, would benefit from mixed-technique/theory approaches. It's good that this fact is brought up in the paper.

Overview of the field. Maybury seems to know what has been done in the past regarding summarization.

Applied to a couple of real-world scenarios. Eventhough there's an imbalance between the efforts spent in the two scenarios presented, it's good that Maybury points to two such different areas of application.

3.2 Cons

Evaluation. A bit weak and unclear when it comes to evaulation; reported work on automatic summarization often acknowledge the difficulties in evaluating the performance of summarizers, but they seldom lend themselves to delve deeper into the issue, rather, the difficulties seem to be used as excuses for *not* doing so. This work is no exception to that trend.

Ad-hoc theory. Mixing techniques is *not* a reason to at least try to be consistent/explanatory when it comes to theory. I'm under the impression that the techniques used for, e.g., selecting and condensing relevant events are somewhat ad-hoc.

Lengthy report. I'm convinced that the contents of the paper could have been presented in less space.

References

Maybury, T. Mark. 1999. Generating summaries from event data. In Mani, Inderjeet and Maybury, Mark T. editors, *Advances in Automatic Text Summarization*, chapter 17, pages 265–281. MIT Press, Cambridge, Massachusetts, London, England.

Sparck Jones, Karen. 1993. What migth be in a summary? In Knorz, G; Krause, J and Womser-Hacker, C editors, *Proceedings from Information Retrieval 93: Von der modellierung zur Anwendung*, pages 9–26.