Summarisation and Novelty in Mobile Information Access: An experimental investigation

Simon Sweeney, Fabio Crestani
University of Strathclyde, Glasgow, UK
{simon, fabioc}@cis.strath.ac.uk

David E. Losada
Universidad de Santiago de Compostela, Spain
dlosada@usc.es

Introduction:

❖ Growing need to deliver information on request, in a form that can be readily and easily digested on the move –*anytime, anywhere information access*, continues to be a challenge.

❖ Automatic text summarisation is a potential solution to achieving device-friendly content for devices that have limited display screens. An effective way to produce a short summary maybe to include only novel information.

❖ The aim was to establish whether a summary that contains only novel sentences provides sufficient basis to determine relevance of a document, or do we need to include additional sentences in the summary to provide context?

Generating Novel Summaries:

❖ We adopted two strategies to produce summaries incorporate novelty in different ways; an incremental summary ($SumN_i$) and a constant length summary ($SumN_c$).

❖ The starting point for generating our novel summaries is an initial seed summary, Sum_1, which is a *query-biased summary* [3].

❖ Given a ranked set of sentences, s_{r1}, s_{r2} , ..., s_{rn} (relevance-based ranking), Sum_1 is composed of the top l_1 sentences ordered as they appear in the original document, where the length is determined as a percentage of the original document length.

❖ Subsequent summaries are generated to include only novel information, and reflect previously seen summary content (similar to *NewWords* in [1]).

- **$SumN_i$** increases the length and increments the size of the next summary to be l_2, = K * l_1, where K = 2, for example, as is the case reported here. This method produces a new summary where all of the material, which appeared in Sum_1 is also present in $SumN_{i2}$.
- **$SumN_c$** maintains a constant length and takes a very different approach producing a new summary, $SumN_{c2}$, whose size l_2 is equal to l_1. Here, we avoid the presentation of material that the user has already seen, and instead focus on the sentences which, in the original (relevance-based) rank, were ranked right after the ones selected for Sum_1.Then, $SumN_{c2}$ is composed of sentences selected from $s_{r(l1+1)}$, $s_{r(l1+2)}$, ..., s_{rn}.

❖ Comparing the 2 approaches, the increasing length method ($SumN_i$) includes both the new sentences and the material already seen, *which we consider as the context*.

❖ On the basis of the score ranking (relevance + novelty) and on the required size, a summary is produced. The top scoring candidate sentences form the final summary. The final stage of the process involves reordering summary sentences according to their ordinal position as they occurred in the original document.

Summaries used in the Experiment:

❖ *Figure 1* shows both the levels and types of summaries prepared, and *Table 1*, example summaries for a sample document.

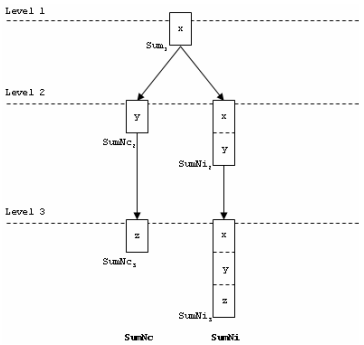


Figure 1: Illustrating the summary types built for the user study.

Level	$SumN_i$	$SumN_c$
1	0,1,5	0,1,5
2	0,1,5,15,16,19	15,16,19
3	0,1,5,7,8,15,16,19,20	7,8,20

Table 1: Listings of summary sentence IDs for summaries of a typical document.

Assumptions:

❖ We restrict the number of summary levels to 3, primarily to avoid overburdening users' in the experimental tasks.

❖ Document titles were included with summaries to assist users in associating summary levels with the source text.

❖ In terms of summary length, for each document a no. of sentences equal to 7% of its length (with a min. of 2 sentences and max. of 6 sentences) were used [2].

Experimental Settings:

❖ The documents were taken from the AQUAINT collection of the TREC Novelty track and consisted of newswire stories (NYT, APW).

❖ A total of 5 randomly selected TREC queries and for each query, the 10 top-ranking documents were used as an input for summary generation.

❖ Experimental Measures: *Precision* (P), *Recall* (R), *Decision Correctness* (DC) and Time.

- DC = (No. of documents correctly marked relevant + No. of documents correctly marked non-relevant) / Total no. of documents marked for that query.

❖ We recruited 20 users to form four experimental groups ($Group_1$ to $Group_4$). Refer to *Table 2*.

❖ Participants were recruited from members of staff and postgraduate students of the Department of CIS at the University of Strathclyde.

Order	Group			
	1	2	3	4
1	$SumB_i$	$SumN_i$	$SumB_c$	$SumN_c$
2	$SumN_c$	$SumB_c$	$SumN_i$	$SumB_i$
3	$SumB_i$	$SumN_i$	$SumB_c$	$SumN_c$
...	...	...	...	...

Table 2: Assignment of summaries to the experimental user groups ($Group_1$: users 1-5; $Group_2$: users 6-10, $Group_3$: users 11-15; and $Group_4$: users 16-20).

Experimental Procedure:

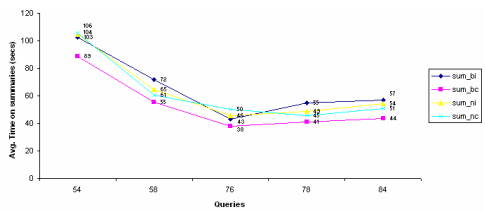
(For details please refer to the paper).

Results:

❖ Results show a slight increase in DC and R performance with $SumN_i$ (summaries that provide novelty with additional context). For P, the baseline* summary with a constant length, $SumB_c$, performs best.

Group	Type	DC	P	R	Time (secs)
1 & 4	$SumB_i$	0.764	0.822	0.845	66
2 & 3	$SumB_c$	0.768	0.850	0.798	53
2 & 3	$SumN_i$	0.776	0.809	0.852	64
1 & 4	$SumN_c$	0.760	0.803	0.752	63

❖ However, the margins of improvement are somewhat minimal. Appropriate statistical tests found no significance difference in the overall results for the different approaches.



❖ Difference in the time spent on $SumN_i$ compared to $SumN_c$ does not agree with what we might normally expect. A possible reason to explain the similarity in viewing times could be that users may skim the longer summaries, glancing over familiar parts, content already seen, and instead focusing on the new parts. The baseline summaries follow a more expected pattern, though again the margin of difference is small.

*We used query-biased summarisation to generate the baseline summaries, and they form the basis of our comparisons.

Conclusion and Future work:

❖ Findings from the user study suggest that there is little difference in performance (DC, P and R) between novel summaries that include context ($SumN_i$) and those that contain only novel information ($SumN_c$).

❖ Therefore, for mobile information access, where issues of bandwidth and screen size are paramount, then we can conclude that an *effective way to produce a short summary is to build one that includes only novel information.*

❖ However, the lack of improvement over the baseline does place doubt over the merit of building novel summaries and will require more investigation.

❖ Extensions to the work we have presented include:

- Investigating the performance of a more refined approach to novelty detection beyond a simple count of new words.
- Study the effects of permitting users to make decisions at any levels; to investigate summary level preference and if there is a corresponding impact on accuracy.

Acknowledgements:
This work is supported by the EU Commission under the IST Project Reveal-This (IST-511689). Information about Reveal-This can be found at http://www.reveal-this.org/.

David E. Losada thanks the support from the "Ramón Cajal" program, whose funds come from "Ministerio de Educación y Ciencia" and the FEDER program. His research is also partially supported by projects TIN2005-08521-01 (from "Ministerio de Educación y Ciencia") and PGIDT03PXIC10501PN (from "Xunta de Galicia").

References:

[1] J. Allan, C. Wade, and A. Bolivar. Retrieval and novelty detection at the sentence level. In *Proceedings of ACM SIGIR'03*, pages 314–321, Toronto, Canada, July 2003.

[2] S. Sweeney and F. Crestani. Effective search results summary size and device screen size: Is there a relationship? *Information Processing and Management*, 42(4):1056–1074, 2006.

[3] S. Sweeney, F. Crestani, and A. Tombros. Mobile Delivery of News using Hierarchically Query-Biased Summaries. In *Proceedings of ACM SAC'02*, pages 634–639, Madrid, Spain, March 2002.

For more information please e-mail *simon@cis.strath.ac.uk* or *fabio.crestani@cis.strath.ac.uk*