Accurate user directed summarization from existing tools

Mark Sanderson
Center for Intelligent Information Retrieval
University of Massachusetts
Amherst, MA, 01003, USA
+1 413 545 3057
sanderso@cs.umass.edu

1. ABSTRACT

This paper describes a set of experimental results produced from the TIPSTER SUMMAC initiative on user directed summaries: document summaries generated in the context of an information need expressed as a query. The summarizer that was evaluated was based on a set of existing statistical techniques that had been applied successfully to the INQUERY retrieval system. The techniques proved to have a wider utility, however, as the summarizer was one of the better performing systems in the SUMMAC evaluation. The design of this summarizer is presented with a range of evaluations: both those provided by SUMMAC as well as a set of preliminary, more informal, evaluations that examined additional aspects of the summaries. Amongst other conclusions, the results reveal that users can judge the relevance of documents from their summary almost as accurately as if they had had access to the document's full text.

1.1 Keywords

User directed summarization, information retrieval, consistency of relevant judgements.

2. INTRODUCTION

The automatic summarization of documents has long been of interest to researchers with work dating back to the fifties where Luhn reported a simple sentence extraction technique based on term frequency [1]. Over the years, increasingly sophisticated methods have been explored, for example, the anaphora resolving summarizer of Paice and Jones [2] or the trainable summarizer of Kupiec [3].

One avenue that has not been addressed until recently is that of *user directed summaries*: summaries that are in some way biased towards the needs of a user. In the context of *Information Retrieval* (IR), an expression of user need is readily available in the form of a text query. In the past two years there has been a growth of interest in the utility of user directed summaries in IR. Tombros and Sanderson [4] conducted a user study investigating the utility of, what they called, *query biased summaries* which were presented in a ranked document list in place of the more usual showing of the first few lines of a document. Their study showed users were better able to identify relevant documents when using the summaries.

Recently, the TIPSTER funded SUMMAC project offered an opportunity for researchers to have their summarization systems evaluated. Participants were asked to summarize a common set of documents and send them to SUMMAC. The accuracy of each participant's summaries was evaluated centrally by a common set of judges. Cross comparisons between participating systems were then possible. SUMMAC addressed a number of summarization tasks, one was concerned with user directed summaries, biased towards an IR query.

This paper first describes the SUMMAC user directed task[1] and its means of evaluation. The design and testing of a summarizer built for the task is outlined next, followed by a presentation of the formal SUMMAC results along with some additional observations regarding the consistency of manual relevance judgements. The paper concludes with a discussion of possible future work.

3. THE SUMMAC EVALUATION

The SUMMAC user directed summary task was defined as follows. Given a query and a set of documents, participants

[1] Only the *fixed-length summary* part of this task is described, as it was the focus of the majority of the author's efforts.

had to automatically generate a summary of each document that would enable a person to judge if the document being summarized was relevant, or non-relevant, to the query. Summaries could be no longer than 10% of the documents' original length and could not contain their title.

The newspaper and newswire part of the TREC IR test collection (composed of the Wall Street Journal, San Jose Mercury News, Associated Press and Zipf newswires) was used as the source of documents. This collection also contained a set of topics, TREC's name for queries, and an accompanying set of relevance judgements, a list of documents that were manually assessed for their relevance, or lack of it, to a topic. From this resource, SUMMAC selected 30 topics and a set of relevant and non-relevant documents for each topic. This was split into a training set, to allow participants to initially test their system, and a test set, from which participants submitted their summaries to SUMMAC for evaluation. The training set was composed of 10 topics, 100 documents per topic, the test set had the remaining 20 topics each with an accompanying 50 documents.

Measurement of summary quality was as follows. A relevance judgement made from a summary was compared to the judgement made by a TREC assessor, who had access to the full text of the document and, therefore, was assumed to be more accurate. In both SUMMAC and TREC, a document was judged relevant if at least one sentence within it was regarded as relevant to the topic. The standard statistics, recall (*R*) and precision (*P*), along with *F* (which is defined in terms of recall and precision) were used to assess the accuracy of the judgements. All three measures are defined below.

$$R = \frac{|r_S \cap r_T|}{|r_T|} \qquad P = \frac{|r_S \cap r_T|}{|r_S|} \qquad F = \frac{2RP}{R+P}$$

The set *r* contains the documents judged relevant by the SUMMAC (*S*) or the TREC (*T*) assessors. The time taken to make relevance judgements was measured as well.

It was anticipated that there would be differences of opinion between TREC and SUMMAC assessors on the question of relevance. Therefore, a number of full text documents were re-assessed by SUMMAC in the same way that summaries were assessed. This established the level of agreement between TREC and SUMMAC assessors and provided an *upper bound* on the experiment. A *lower bound* was supposed to be established also by having a number of assessors judge summaries consisting of the first 10% of a document. As these were news articles, the opening few sentences were assumed to be an author-generated summary of the article. Due to unknown circumstances, however, this lower bound or baseline was not properly established by SUMMAC.

It was in this experimental environment that the design, building and initial testing of the summarizer took place.

4. DESIGN OF THE SUMMARIZER

The primary aim in the design of the summarizer was to base it on well established statistical techniques taken from IR. The reasons behind this were two fold:

- Firstly, it would establish a simple 'statistical baseline' against which more sophisticated linguistically based methods could be compared against.
- Secondly, by basing it on existing techniques, construction of the summarizer would be relatively simple and fast. The research retrieval system, INQUERY [5], was available for use and it already had the functionality required for the summarizer design. These parts were the *best passage operator* and *Local Context Analysis* (LCA).

4.1 Best Passage

Given a document and a topic, the best passage operator returns the region of the document that best matches the query. This process is performed as a retrieval operation (using stop word removal, stemming, term weighting, etc.), where a set of candidate passages within the document are ranked by the degree to which they match the topic. The top ranked passage contains the highest density of topic terms in the document. Candidate passages are a set of equally sized overlapping regions ranging over the whole of the document. For example, if a passage 100 words in length is required, candidate passages are created over the document from word 1 to 100, from 51 to 150, 101 to 200, 151 to 250, etc. Note that document structures such a paragraph and sentence boundaries are ignored. This operator has been used successfully as a means of improving retrieval effectiveness; it is described in more detail by Callan [6].

In the context of user directed summaries, a single 10%-sized best passage of a document was used as its summary. Other options were considered, such as two 5% sized passages, however, it was felt that a single coherent passage would be more readable and so was chosen.

4.2 LCA

Local Context Analysis is a form of automatic query expansion using a so-called *pseudo feedback* technique which works as follows. Given a topic and a document collection, the LCA process first performs a retrieval based on the topic and selects the top ranked documents (e.g. top 50) resulting from that retrieval. It then examines in each document the context surrounding the topic terms. Words/phrases that are unusually frequent in these contexts (in comparison to their frequency of occurrence in the document collection as a whole) are selected and added to the query. For example, given the topic

> Reporting on possibility of and search for extra-terrestrial life/intelligence.

The top 70 words/phrases from LCA (processing the part of the TREC collection SUMMAC documents were drawn from) are as follows.

> extraterrestrials, planetary society, universe, civilization, planet, radio signal, seti, sagan, search, earth, extraterrestrial intelligence, alien, astronomer, star, radio receiver, nasa, earthlings, e.t., galaxy, life, intelligence, meta receiver, radio search, discovery, northern hemisphere, national aeronautics, jet propulsion laboratory, soup, space, radio frequency, radio wave, klein, receiver, comet, steven spielberg, telescope, scientist, signal, mars, moises bermudez, extra terrestrial, harvard university, water hole, space administration, message, creature, astronomer carl sagan, intelligent life, meta ii, radioastronomy, meta project, cosmos, argentina, trillions, raul colomb, ufos, meta, evidence, ames research center, california institute, history, hydrogen atom, columbus discovery, hypothesis, third kind, institute, mop, chance, film, signs

Xu and Croft [7] demonstrated the utility of LCA as a means of improving the effectiveness of a retrieval system.

In the context of the summarisation task, LCA was used to expand topics with an additional 70 words/phrases. The expansion was based on the newspaper and newswire part of TREC.

The INQUERY based summarizer used in the SUMMAC evaluation, therefore, was a system that found the best passage in a document based on an LCA expanded topic. In the unlikely event of there being no topic or LCA terms found in the document, the summarizer produced a message instructing the assessor to regard the document as not relevant.

5. PRELIMINARY TESTING

Prior to the main SUMMAC evaluation, the released training set of topics, documents, and relevance judgements was used to measure the summarizer's accuracy and test it in an alternate configuration. The number of experiments was kept to a minimum as measurement was a time consuming process requiring assessors to read hundreds of summaries and judge their relevance. In this exercise, two assessors performed relevance judgements: the author and one other. Each assessed half the topics supplied with the training set, 100 summaries were generated per topic. When measuring the summarizer in a new configuration, the assessors were given the same topics to process. It was assumed that the assessors would not change their view of what constituted a relevant document across the assessments. Although this is far from an ideal experimental set up, it was expected that large differences in summary quality would still be observable.

	R	P	F
Summarizer	69	68	68

Table 1: Measurement of summarizer accuracy in its default configuration.

	R	P	F
Summarizer	69	68	68
Lower Bound	32	71	44

Table 2: Comparison of lower bound with summarizer accuracy.

5.1 Summarizer accuracy

In the first experiment, the summarizer was tested in its default configuration of using LCA to produce a single 10% sized best passage. The results of this experiment are shown in Table 1. They were regarded as being somewhat disappointing but without the establishment of a lower or upper bound, it was impossible to interpret these results. No formal upper bound was established in the training set, though an informal check of the differences between the experiment's assessors and the TREC judgements revealed much disagreement on what constituted a relevant document. This hinted that the upper bound of accuracy might not be all that different from that measured for the summarizer. As will be shown in Section 6, this was a reasonably correct supposition.

5.2 Measuring the Lower Bound

As all documents in the SUMMAC corpus were news articles, it was presumed that the opening section of each article was a summary. Since the author (journalist or sub-editor for these articles) wrote the summary, and most news stories cover a single topic, the summary could be expected to provide a well thought out precis that would provide a user with good evidence on the relevance of a document to a topic.

In this evaluation of the summaries, the two assessors were presented with the first 10% of each document. The resulting accuracy of their relevance judgements is shown in Table 2 along with the accuracy of the summarizer. As can be seen, there is a large reduction in accuracy indicating that the author summaries were not as good as INQUERY's. An examination of relevant documents that were missed by the assessors revealed that in these cases the topic of interest played only a small part in the document's story and, therefore, was not mentioned in the opening paragraph. For example, the topic

> What is the main function of the Federal Emergency Management Agency (FEMA) and the funding level provided to meet emergencies? Also, what resources are available to FEMA such as people, equipment, facilities?

and document "SJMN91-06303216": the baseline summary (shown below) provided no information regarding FEMA;

> Aerial seeding finally got under way in the fire-ravaged Oakland hills Monday as authorities continued their search for the cause of the massive Oct. 20 fire that claimed the lives of 25 people.; A helicopter that had been waylaid first by mechanical problems and then stiff wind finally made it into the air above the fire-devastated hills to begin reseeding the 1,400 blackened acres. "We did about 40 percent of the work," said Surlene Grant, spokeswoman for the Oakland…

the INQUERY summary (taken from further down the document) revealed, however, that the relevant section was only a side issue of the story.

> …Emergency housing assistance checks Monday.; Just eight days after the Oakland hills disaster, FEMA was ready with $47,700 in checks for those who had no insurance or whose policies did not provide them with emergency rental assistance.; Normally checks are mailed, and they will be beginning today.; "It was to demonstrate that we are here and that we are getting them (checks) out as fast as we can," said Susan Robinson, disaster assistance employee with FEMA.; Chuck…

One could draw two conclusions from this result.

First, even though it was worse than the summarizer was, the experiment showed that the opening 10% of a document did provide a reasonable indication of relevance for a significant fraction of documents[2].

The second conclusion is that the result supported the view that what the author thought was the main point of the article was less important than user information needs. One could sum up this conclusion by saying

> The author wrote the article, but he/she does not know why it is going to be relevant.

5.3 Building Summaries Without LCA

Although LCA had been shown to work well for improving IR, it was unclear how it well it would work in aiding summarization. Therefore, an experiment was conducted where the selection of a best passage (i.e. summary) from a document was based on the original topic terms alone. The results of this evaluation along with previous ones are shown in Table 3. As can be seen, there was effectively no

	R	P	F
With LCA	69	68	68
Without LCA	67	72	70
Lower Bound	32	71	44

Table 3: Comparison of summarizer in two configurations: with and without use of LCA expansion terms.

difference in the assessor's accuracy of relevance judgement between the two runs. This indicated that LCA provided nothing in terms of locating the relevant part of a document. From this result, it would appear that all an assessor required to judge document relevance was to read how some of the original topic terms were used within the document. Indeed, it may be more realistic to abandon the notion that summaries are being generated and to simply recognise that all that is important for IR is to show users text from the document to enable them to better understand how their query words are used in that document.

It may be, however, that some aspect of the training set caused LCA's impotence. A strong possibility is the length of the topics. Like most generated from TREC, they were long, on average 38 words in length (after stop words were removed). It may well be that LCA would have been of more use when working with the much shorter queries (2-3 words in length) often entered by casual users.

One further aspect of this somewhat unexpected experimental result was that it was thought that LCA expansion might be able to generate a user directed summary of a document which contained none of the original topic terms. It was found that approximately 11% of the documents processed were in this state and LCA expansion terms had indeed been needed to find a best passage. However, all but one of the documents was not relevant. The single example of a relevant document that contained no topic terms is shown here. For the topic (shown in Section 4.2),

> Reporting on possibility of and search for extra-terrestrial life/ intelligence[3].

the following summary of the document, "SJMN91-06220286", was produced (matching LCA terms are highlighted).

> ...up," he said.; So why would the government want to cover something up? Friedman has his share of theories.; For one thing, if the U.S. government got hold of a highly sophisticated **alien** spacecraft, it might think twice about sharing that information with the rest of the world. "If you tell your friends, you tell your enemies," said Friedman.; As one might expect,

[2] This conclusion may apply to other passages of a document, as it is possible that any randomly chosen passage of similar length could provide an equally good (conceivably even a better) indication of a document's relevance. To establish the true utility of the leading passage, it would be necessary to measure the accuracy of relevance judgements based on passages randomly chosen from documents. This undertaking is left for others to pursue, however.

[3] There actually was one topic term elsewhere in this document, "extraterrestrial", however, due to a quirk in the way INQUERY processed hyphenated words, it was not matched.

> Friedman doesn't have much patience for "noisy negativists," including former classmate **Carl Sagan**. They proclaim that **UFOs** -- like the tooth fairy -- just plain don't exist. One critic has accused Friedman of "making a buck off people's credulity.";…

Despite the apparent lack of utility of LCA, the configuration of the summarizer used to process the SUMMAC test data was left with LCA in. It was felt that at least the technique was not harming accuracy and may yet prove its worth on other situations.

6. SUMMAC TEST RESULTS

A total of eleven groups participated in the fixed length user directed summary task. The results, taken from the SUMMAC assessment [8], are shown in Table 4 sorted by *F* score. As can be seen, there are marginal differences between the submitting groups. Significance tests conducted by SUMMAC partitioned the participants into three groups with the INQUERY summarizer in the middle group, ranked third overall.

The top three summarizers (CGI/CMU, Cornell and INQUERY) all employed statistical techniques using a form of paragraph/passage extraction based on similarity with the topic. However, both Cornell and CGI/CMU's system were more sophisticated than INQUERY in their passage extraction. When selecting passages, both systems tried to find ones that were highly ranked and different from each other (as measured by some statistical word overlap measure). Presumably this was an attempt to provide a summary with a wider coverage of the document.

The Cornell summarizer (apparently based on work described in Salton et al [9]) selected the three highest ranked 'different' paragraphs of a document and from this extracted their sentences. These were then ranked based on their similarity to the topic and the top ranked were selected and presented as a summary.

CGI/CMU used a similar technique though they extracted sentences directly from the document and ranked them by their similarity to the topic. Like Cornell, they used a statistical measure (called the Maximal Marginal Relevance Metric) which selected top ranked, yet different, sentences. In addition, the CGI/CMU summarizer gave strong preference to including a document's opening sentence in the summary. As indicated in the lower bound experiment of Section 5.2, this may have played a part in the superior accuracy of their summaries.

Although it was not the best, the INQUERY system appeared to be one of the top performing systems in the evaluation. Given its lack of regard for textual structure and its simplistic passage selection methods, performing so well in the assessment was unexpected. If there was any conclusion to be drawn about the system from this exercise, it was that the selection of a single top ranked passage was most likely to be the cause of it doing worse than Cornell and CGI/CMU. Although SUMMAC results revealed that INQUERY summaries were read slightly quicker than the top two systems, it is probably preferable to sacrifice ease of reading for judgement accuracy.

Run	P (%)	R (%)	F (%)
Upper bound	84	63	72
CGI/CMU	76	52	62
Cornell	78	47	59
INQUERY	79	47	59
Group 7	78	46	58
" **1**	81	45	57
" **16**	76	46	57
" **8**	80	42	55
" **4**	78	41	54
" **2**	78	36	50
" **9**	78	37	50
" **15**	83	34	48

Table 4: Precision, Recall and *F* scores of groups in SUMMAC fixed length adhoc task. Thin lines indicate a statistically significant difference between the groups above and below the line.

6.1 Quality of the Summaries

One of the strongest results from the SUMMAC evaluation was that the 10% summaries allowed relevance assessors to make assessments almost as accurately as if they had had access to the full document text. With the INQUERY based summarizer, for example, assessors identified 75% of the relevant documents they would have identified had they had access to the full text[4]. In addition to this, assessors took on average 24 seconds to assess each summary whereas they took 61 seconds to assess the full documents. This 60% reduction in time taken indicates that the 10% fixed length summaries work well for the relevance assessment task.

6.2 Conclusions From SUMMAC

An overall conclusion at SUMMAC was that the user directed summary task was too easy. Relatively simplistic summarizers like that described in this paper generated summaries of such high quality that there appeared to be little room for improvement. It may be well that the reasons for this success was due to the type of relevance judgements made in TREC where a document is judged relevant if just one of its sentences is considered relevant to the topic. In an alternative corpus or when relevance is defined differently, the summaries may be of less use. However, if

[4] This figure was calculated by dividing the recall value of INQUERY against that of the upper bound, see Table 4.

a user directed evaluation like the SUMMAC task were to be repeated it would seem prudent to at least change the corpus being experimented upon.

7. Does Accuracy Vary Over a Ranking?

Given that one would expect user directed summaries to appear in the context of an IR system, it seemed reasonable to examine if summaries of top ranked documents might be easier to judge than summaries of lower ranked documents. In order to evaluate this, a retrieval for each of the 20 test set topics was performed on a document collection containing the test set documents that were summarised. The retrieval system used was INQUERY. The rank position of the documents was noted and the SUMMAC provided *F* measures of summary accuracy were calculated for documents ranked as positions 5, 10, 20, 50, 100, 250, 500 and 1,000. Table 5 shows the result of this measurement, which shows a higher *F* score for higher ranked documents. This result appeared to imply that assessors were able to make more accurate relevance judgements on higher ranked documents.

On closer inspection, the result turned out to be merely an artifact of the measuring process. For it to be true, the level of agreement between SUMMAC and TREC relevance assessments on full text documents (i.e. the upper bound measurement) would have had to have been the same no matter where the documents being assessed were ranked. Such consistency between assessors did not occur: using the same method as in the experiment above, the SUMMAC and TREC assessor agreements (on the relevance of full text documents) was measured in the context of a ranking. It revealed a higher level of agreement between assessors for top ranked documents and a lower level of agreement for low ranked documents. Table 6 shows that this variation is the same as that found in Table 5 and, therefore, the variation of summary accuracy, shown above, was concluded to be due to the variation of assessor agreement only.

Though Lesk and Salton [10] first described this quality of relevance assessments in 1969, it is believed that the experimental result just described constitutes the first time this quality has been demonstrated over a relatively large data set (i.e. 1,000 relevance judgements). One can draw a number of conclusions from this result.

First, it indicates that IR systems seem to rank highest those documents that people agree are relevant. Documents where there is disagreement are more likely to be ranked lower. This would appear to be a good thing. Second, given that top ranked documents generally contain a larger number of topic terms within them than lower ranked documents, it would indicate that people are more likely to agree that a document is relevant if it has more query terms within it.

This is an area that would most likely benefit from further investigation.

Rank	5	10	20	50	100	250	500	1000
R (%)	64	62	55	56	53	54	50	47
P (%)	78	81	80	79	76	76	75	79
F (%)	70	70	65	66	62	63	60	59

Table 5: Changes in summary accuracy at different rank positions.

Rank	5	10	20	50	100	250	500	1000
R (%)	80	78	76	70	67	64	63	59
P (%)	84	84	81	80	77	77	78	82
F (%)	82	80	78	75	72	70	69	68

Table 6: Changes in inter-assessor agreement at different rank positions.

8. FUTURE WORK

As has already been stated, the user directed summary task of SUMMAC was too easy. Any future research would probably benefit from addressing different issues such as alternative summary types or other corpora.

As indicated in Section 6, INQUERY's selection of a single top ranked passage was its most likely failing. In general, the SUMMAC corpus consisted of relatively short documents (~1,000 words in length) for which a single passage summary was quite readable. However, for the occasional long document the passage was too large and took too long to read. A summarisation method that in these cases selected a number of passages may have worked better. Indeed, the issue of summary presentation was not addressed at all in the evaluations and this also may be a useful line of inquiry.

Given the relative success of the summarisation system in allowing assessors to judge the relevance of documents quickly and accurately, it may be worth investigating the use of a summarisation system when generating relevance assessments for a document test collection. Although the assessments would be less accurate than if an assessor had seen the full document text, because assessors would process the summaries in a much shorter time, more relevance judgements would be made. Such a strategy could prove to be of benefit in the creation of test collections.

9. CONCLUSIONS

In this paper a statistically based user directed document summarizer was described and its evaluation reported. The design of the summarizer was based on existing components of the INQUERY retrieval system: namely its best passage operator and query expansion from Local Context Analysis (LCA).

The initial evaluation on the SUMMAC training corpus revealed that it was summarizing better than a lower bound

summarizer was but that use of LCA expansion terms was providing no improvement to the summary quality. In the full SUMMAC evaluation, the summarizer was one of the better performing systems, demonstrating that the summaries generated allowed users to quickly and accurately assess the relevance of documents. The final part of the evaluation addressed the issue of summary quality in terms of the rank position of the summarised documents. Here it was found that relevance assessors were in more agreement over the relevance of top ranked documents than they were about lower ranked documents.

10. ACKNOWLEDGMENTS

This material is based on work supported in part by the National Science Foundation, Library of Congress and Department of Commerce under cooperative agreement number EEC-9209623. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor.

Thanks to Jamie Callan, Pam Sanderson and the anonymous reviewers (especially the one who wrote a page and a half of comments) for their advice, help and comments on this work.

11. REFERENCES

[1] Luhn, H.P. The automatic creation of literature abstracts: IBM Journal, pp. 159-165, 1958

[2] Paice, C.D., and Jones, P.A. The identification of important concepts in highly structured technical papers: In the proceedings of the 16th ACM SIGIR conference on research and development in information retrieval, pp. 69-77, 1993.

[3] Kupiec, J., Pedersen, J.O., and Chen, F. A Trainable Document Summarizer: In the proceedings of the 18th ACM SIGIR conference on research and development in information retrieval, pp. 68-73, 1995.

[4] Tombros, A., and Sanderson, M. Advantages of query-based summaries in IR: To appear in the Proceedings of the 21st ACM SIGIR conference, 1998.

[5] Allan, J., Callan, J., Croft, W.B., Ballesteros, L., Byrd, D., Swan, R., and Xu, J, INQUERY does battle with TREC-6: In the 6th Text REtrieval Conference (TREC-6), NIST special publication.

[6] Callan, J., Passage-Level Evidence in Document Retrieval: In the proceedings of the 17th ACM SIGIR conference on research and development in information retrieval, 1994.

[7] Xu, J., and Croft, W.B. Query expansion using local and global document analysis: In the proceedings of the 19th ACM SIGIR conference on research and development in information retrieval, pp. 4-11, 1996.

[8] Firmin, T., and Sundheim, B. TIPSTER/SUMMAC summarization analysis participant results: Presented at the TIPSTER text phase III 18-month workshop, 1998. (http://www.tipster.org/).

[9] Salton, G., Allan, J., Buckley, C., and Singhal, A. Automatic analysis, theme generation, and summarization of machine-readable texts. Science, 264, 1421-1426, 1994.

[10] Lesk, M.E., and Salton, G. Relevance assessments and retrieval system evaluation: in Information Storage and Retrieval, Vol. 4 pp. 343-359, 1969.