Total Recall: Blue Sky on Mars

Charles L. A. Clarke, Gordon V. Cormack, Jimmy Lin, and Adam Roegiest
David R. Cheriton School of Computer Science
University of Waterloo, Ontario, Canada
claclark@gmail.com, {gvcormac, jimmylin, aroegies}@uwaterloo.ca

ABSTRACT

There are presently plans to create permanent colonies on Mars so that humanity will have a second home. These colonists will need search, email, entertainment, and indeed most services provided on the modern web. The primary challenge is network latencies, since the two planets are anywhere from 4 to 24 light minutes apart. A recent article sketches out how we might develop search technologies for Mars based on physically transporting a cache of the web to Mars, to which updates are applied via predictive models. Within this general framework, we explore the problem of high-recall retrieval, such as conducting a scientific survey. We explore simple techniques for masking speed-of-light delays and find that "priming" the search process with a small Martian cache is sufficient to mask a moderate amount of network latency. Simulation experiments show that it is possible to engineer high-recall search from Mars to be quite similar to the experience on Earth.

1. INTRODUCTION

Mars needs search. And Mars needs recommendation, social media, streaming video, email, messaging, e-commerce, and indeed most services provided on the modern web. Unfortunately, providing these services for our permanent Martian colonists poses non-trivial challenges, most particularly due to network latencies associated with users who are 4 to 24 light minutes away. On Earth, a query to a commercial search engine takes under a second to return results. On Mars, without appropriate technology, the same query would take from hundreds to thousands of times longer.

Plans for permanent Martian colonies continue to move forward (see marstostay.com and mars-one.com), with public support from prominent figures such as Apollo astronaut Buzz Aldrin [1] and entrepreneur Elon Musk [7]. These plans call for colonists to move permanently to Mars, with no expectation of return. While some colonists may eventually return to Earth, as Martian resources and industry permit, some colonists may live their entire lives there. They will start families, grow old, and die there. In the short term the colonists would conduct scientific missions. In the longer term they may take on larger projects, including terraforming the planet. Humanity would have a second home.

Permanent Martian colonies are economically feasible with current science, requiring no new breakthroughs. However, making colonization a reality will require a substantial effort to create the individual technologies needed to support life on Mars, including technologies for information access. To conduct scientific work, but also to maintain important social and cultural connections to Earth, searching the web should be as easy from Mars as it is from Marseille.

In a recent article, Lin et al. [10] examined the requirements for searching from Mars. They envision the Martians of the future using the web much like the Earthlings of today: reading, watching, gaming, and interacting—even buying physical items as gifts for Earth-bound friends and family, or as precious cargo for themselves, delivered on the next supply freighter. These tasks must be accomplished despite round-trip latencies ranging from 8 to 48 minutes, depending on the relative positions of the two planets. With laser-based communication [11], reasonable bandwidth is possible, but physical laws prohibit latency improvements.

Lin et al. proposed a solution for searching on Mars that starts with a physical copy of the web shipped to Mars as cargo (i.e., the first interplanetary sneakernet), arriving either with or before the first colonists. Upon arriving on Mars, it becomes a cache for our Martian search engine. By the time it arrives, however, months or years later after liftoff from Earth, this cache will be stale, and so ongoing updates must be applied with live Earth data. Since bandwidth will remain precious for the foreseeable future, and we assume that all resources are to be used as parsimoniously as possible, these updates must be performed intelligently. In order to provide Earth-like response times, the Martian search engine must anticipate future searches and other interactions, prefetching data as needed using predictive models on Earth that act as proxies for the colonists on Mars. Lin et al. posed a challenge to information retrieval researchers to develop and validate these models.

In this paper, we take a small step towards our goal of searching on Mars. We focus on a single problem, that of high-recall retrieval, and examine the impact of high latency on this problem. We imagine a Martian conducting high-recall retrieval, such as an inquiry into the best mechanisms to achieve potato growth in recycled organic waste or a scientific survey of the flora and fauna of Barsoom. We assume the Martian cache contains some limited information on the

ICTIR '16, September 12 - 16, 2016, Newark, DE, USA

ISBN 978-1-4503-4497-5/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2970398.2970430

retrieval topic, enough to initiate the search process. As the Martian searches, relevance information heads back to Earth, which responds with a stream of potentially relevant documents. We explore the impact of Mars-Earth latency and simple techniques for masking speed-of-light delays via simulations, by comparing against Earth-based search where latency is negligible. We find that "priming" the search process with a local (Martian) cache is sufficient to mask a moderate amount of network latency. Based on our simple techniques, the experience of searching from Mars can be engineered to be similar to searching from Earth.

The contribution of this paper is the first experimental study of searching from Mars, building on the proposals of Lin et al. [10] (who performed no actual experiments). Our work takes a rather (currently) fanciful problem, shows that there are substantive research questions worth exploring, demonstrates the inadequacies of naïvely applying current techniques, and evaluates simple solutions that address the relevant challenges. We hold this paper as an exemplar of how such research can be performed, with potential applications closer to home (e.g., searching from the Canadian Arctic or rural communities in India, both of which suffer from poor connectivity). Teevan et al. [15] proposed "slow search", which aims to relax latency requirements for a potentially higher-quality search experience; search from Mars was mentioned in passing as an illustrative example, but they did not propose any specific solutions.

2. TOTAL RECALL

Given that the focus of this paper is high-recall retrieval, our experiments are based on data from the TREC Total Recall Track [6], which evaluates high-recall retrieval systems using topics drawn from legal and other domains. The goal of the track is to develop and validate methods for identifying all relevant material on a given topic with as little user effort as possible, returning all relevant documents before all non-relevant documents. The track is focused around a task, similar to active learning, in which a (simulated) user judges documents proposed by the retrieval system. Starting with an initial query, the retrieval system aims to return a stream of relevant documents to the user.

2.1 AutoTAR

To facilitate system development for the TREC Total Recall Track [6], the coordinators released an implementation of Cormack and Grossman's AutoTAR [3, 5] protocol, which is itself an extension of their continuous active learning protocol [2, 4]. Since AutoTAR's effectiveness ranks among the best systems in the TREC evaluations, we use it as the basis of our experiments.

AutoTAR initially ranks the entire document collection by training on an initial query (treated as a pseudo-document) and 100 randomly selected documents, which are assumed to be non-relevant. AutoTAR requests that the user assesses the most relevant (i.e., highest scoring) document, and then trains on this new assessment along with the query and 100 new random (assumed non-relevant) documents. In an ideal situation, AutoTAR would repeat this process, selecting the top document for assessment and training on all available assessments. Since this is not computationally practical, AutoTAR requests assessments in exponentially increasing batches, starting with a batch size of one and increasing the batch size by $\textsc{Max}(1, 0.1 \cdot \texttt{batch\_size})$ each iteration.

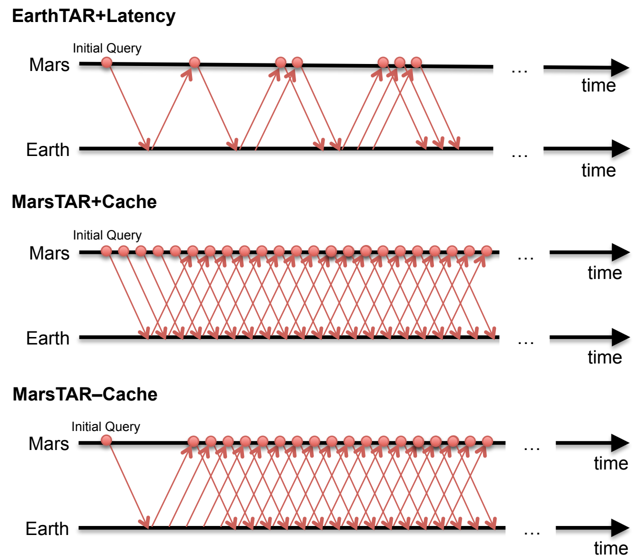

Figure 1: Illustration of various AutoTAR on Mars scenarios. Circles on the Mars timelines indicate relevance judgments.

It has been observed that for high-recall retrieval algorithms [12], assessor background and assessor experience can affect the judgments rendered [8, 17, 14]. Furthermore, changing the training and evaluative assessments can affect system performance in non-trivial ways [13, 16, 18]. In spite of these issues, we follow the Total Recall Track's premise and assume that the training and evaluative assessors are the same (i.e., the Martian researcher) and that the assessments would not change depending on the algorithm employed (i.e., that the Martian would always judge documents the same way, regardless of the order in which they are received).

2.2 AutoTAR on Mars

What would AutoTAR on Mars look like? We assume the same general setup of the Total Recall Track—the user provides binary relevance feedback on each document as it is judged, and the retrieval system uses this feedback to update its notion of relevance. The retrieval process continues back and forth, with the system proposing one or more documents, and the user providing judgments. Except that on Mars, each back-and-forth is subjected to speed-of-light latency constraints, a round trip time of 8 to 48 minutes. We consider the following three scenarios, with varying levels of latency awareness:

EarthTAR+Latency. As a naïve baseline, we could just run a replica of the Total Recall task with the added delay of communications latency, as if the Martian used the Earth-based software with no accommodations. Under this scenario, the Earth-based retrieval system waits for feedback from Mars before proposing new documents. While waiting for results, the Martian could perform other work, e.g., investigating crystalline structures in the Cydonia region of Mars, spreading pesticides to kill off Martian insects in algae fields, etc. This scenario is shown at the top of Figure 1.

MarsTAR+Cache. As an attempt to hide latency, we can stream documents to Mars at the rate they can be judged, with delayed relevance information modifying the stream as it is received. We assume that AutoTAR runs initially over the Martian cache, but once data begins arriving from Earth, this data is added to the cache. Under this approach, two versions of AutoTAR run concurrently: one on Mars with partial knowledge, and one on Earth with access to the complete document collection. The Earth-side AutoTAR sends documents to Mars for incorporation into the local Martian collection when it decides that it has received enough assessments to generate a new batch of documents (using exactly the AutoTAR algorithm described in Section 2.1).

Note that from the Martian perspective, although feedback is not reflected in the stream until after a round-trip communications delay, the searcher perceives no latency—she can continue judging documents in the Martian cache. In essence, we are using the Martian cache to hide latency. This scenario is shown in the middle of Figure 1.

MarsTAR−Cache. What if there are no cached documents on Mars? The Martian begins with the initial query that is sent to Earth, but needs to wait for one round-trip communications delay to have subsequent documents to judge. After that, the searcher can continuously assess documents with a (perceived) latency-free experience. Note that these documents are always the ones relayed from Earth. This scenario is shown at the bottom of Figure 1.

As an upper bound, we consider simply running AutoTAR on Earth, where the latency is negligible—this would be as if the Martian returned to Earth to perform the search, perhaps using Dr. Manhattan-like teleportation. We call this the **EarthTAR** condition, which of course, provides an upper bound on effectiveness. At the other end of the spectrum, EarthTAR+Latency serves as a lower bound. The question is, how good are our MarsTAR configurations in being able to replicate the search experience on Earth?

3. EXPERIMENTS AND RESULTS

To determine if high-recall retrieval is possible across intrasolar distances, we conducted a set of simulation experiments using the Reuters Corpus Volume 1 v2 (RCV1) [9], consisting of just over 800,000 newswire documents manually labeled with 103 topic codes. The benefit of using RCV1 is the presence of a time-delimited training and test split of the corpus, which facilitates the accurate simulation of the stale-cache scenario proposed by Lin et al. [10]. The “training” portion (i.e., the Mars cache) consists of the oldest 24,000 newswire documents, with the “test” portion comprising the remainder of the collection.

We used each of the 103 topic codes as information needs—i.e., the searcher’s goal is to find all documents with a particular assigned code. Topic codes are hierarchically arranged in RCV1: each topic’s pseudo-document (i.e., the initial query) is the topic code (string) label plus all parent labels concatenated together.

To simplify the simulations we assume that the time to send a document from Earth to Mars, a judgment from Mars to Earth, and the time required to judge a document all take one unit of Martian time, which we call a ‘tal’ (following Edgar Rice Burroughs). Since we anticipate that bandwidth to Mars will remain a precious resource, there is no sense to sending documents at a rate faster than the Martian can judge them—in this strategy, we conserve bandwidth for other purposes. Although this temporal coupling is not necessarily realistic, we believe that it is sufficient for the scenario we are trying to simulate.

Accordingly, we have chosen four levels of round-trip latencies between Earth and Mars: 30, 100, 300, and 1000 tals. We evaluate these variants of AutoTAR by measuring recall as a function of tals elapsed from the initial query time on Mars. To summarize the results of all 103 topics, we average recall across topics at fixed points of tals elapsed in simulation time (e.g., 100, 200, etc.) to generate a gain curve.

Results of our simulations are shown in Figure 2. Even with a latency of 30 tals, round-trip time has a clear impact on the EarthTAR+Latency scenario. With a latency of 1000 tals, recall for EarthTAR+Latency remains close to zero out to 5000 tals and beyond. Clearly, naïvely applying Earth-based technology for Mars is hopeless.

Under the MarsTAR+Cache scenario, the gain curves at low latencies approach that of Earth. With longer latencies, the impact of latency is noticeable, but gain begins to recover after the initial round trip. This means that we can replicate a high-recall search experience that is quite similar to search on Earth by masking the effects of latency with a local cache that is *only* 3% of the entire collection. While waiting for communications delay, the Martian can continue refining the system’s model of relevance such that, under relatively short delays, the gain curve is practically indistinguishable from an Earth-based searcher.

Even without a local cache (i.e., MarsTAR−Cache), we find that the search experience degrades negligibly with low latencies. However, with large latencies, effectiveness suffers noticeably. By simply pipelining the relevance judgments, we can mask communications latencies to a large extent. The bottom line: high-recall retrieval from Mars? Totally doable. Pluto, on the other hand, might be a bit trickier.

4. CONCLUSION

I just had a terrible thought... what if this is a dream? — Douglas Quaid

Our line of research requires the acceptance of a fairly major assumption: searching from Mars. While this assumption might be viewed as unlikely in the short term, it is considerably more likely than the zombie apocalypse preparations advocated by the Centers for Disease Control.[1] Like that effort, theoretical considerations about unlikely scenarios can lead to insights with more immediate impact. For example, a better understanding of searching from Mars might lead to improved search from remote areas on Earth, such as Easter Island, where only satellite internet is available, the Canadian Arctic, where internet access remains prohibitively slow and expensive, and even rural villages in India.

Our experiments illustrate a methodology for developing search technologies for Mars. Starting with a defined task, we consider the impact of latency and develop methods to compensate for it. Effectiveness is measured by comparing the task without compensation as a baseline and the task on Earth as the “gold standard” target. We provide an exemplar of this approach in action, applied to high-recall retrieval on Mars, and demonstrate how we can come close to replicating the search experience on Earth.

Acknowledgments. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. Any opinions, findings, conclusions, or recommendations expressed are solely those of the authors.

[1] www.cdc.gov/phpr/zombies.htm

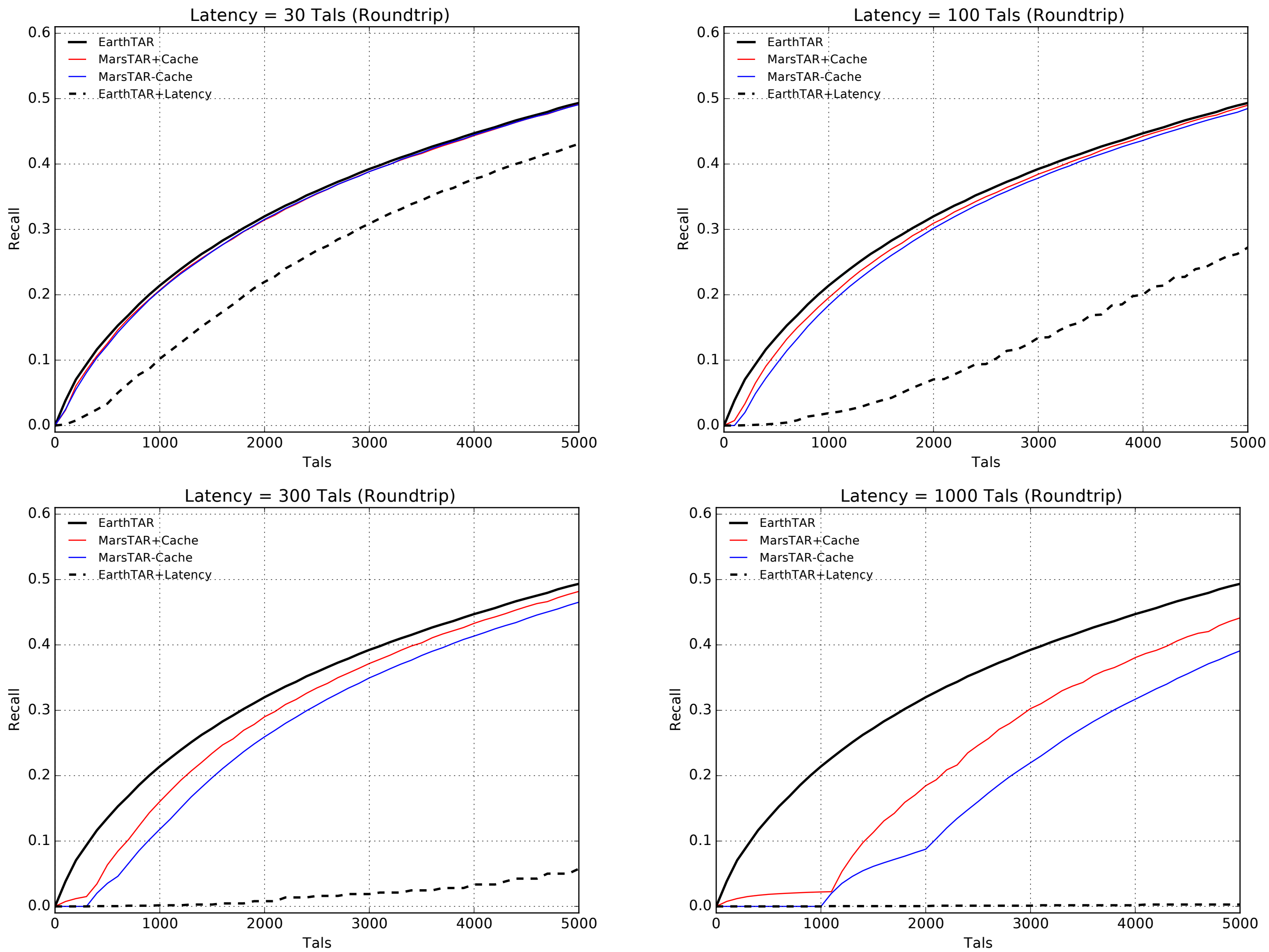

Figure 2: The effectiveness of our proposed strategies for high-recall retrieval from Mars under various round-trip latencies ('tals'). EarthTAR represents the "gold standard" of running AutoTAR on Earth; EarthTAR+Latency represents a naïve baseline with no latency masking. MarsTAR conditions represent pipelining judgments, with and without a local Martian document cache.

5. REFERENCES

[1] B. Aldrin. The call of Mars. *New York Times*, June 2013.

[2] G. V. Cormack and M. R. Grossman. Evaluation of machine-learning protocols for technology-assisted review in electronic discovery. *SIGIR*, 2014.

[3] G. V. Cormack and M. R. Grossman. Autonomy and reliability of continuous active learning for technology-assisted review. *arXiv:1504.06868v1*, 2015.

[4] G. V. Cormack and M. R. Grossman. Multi-faceted recall of continuous active learning for technology-assisted review. *SIGIR*, 2015.

[5] G. V. Cormack and M. R. Grossman. Engineering quality and reliability in technology-assisted review. *SIGIR*, 2016.

[6] G. V. Cormack, M. R. Grossman, A. Roegiest, and C. L. A. Clarke. Overview of the TREC 2015 Total Recall Track. *TREC*, 2015.

[7] E. Howell. SpaceX's Elon Musk to reveal Mars colonization ideas this year. *Space.com*, January 2015.

[8] K. A. Kinney, S. B. Huffman, and J. Zhai. How evaluator domain expertise affects search result relevance judgments. *CIKM*, 2008.

[9] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5, 2004.

[10] J. Lin, C. L. A. Clarke, and G. Baruah. Searching from Mars. *Internet Computing*, 20(1):77–82, 2016.

[11] B. S. Robinson, D. M. Boroson, D. A. Burianek, and D. V. Murphy. The Lunar Laser Communications Demonstration. *ICSOS*, 2011.

[12] A. Roegiest and G. V. Cormack. Impact of review set selection on human assessment for text classification. *SIGIR*, 2015.

[13] A. Roegiest, G. V. Cormack, C. L. Clarke, and M. R. Grossman. Impact of surrogate assessments on high-recall retrieval. *SIGIR*, 2015.

[14] F. Scholer, D. Kelly, W.-C. Wu, H. S. Lee, and W. Webber. The effect of threshold priming and need for cognition on relevance calibration and assessment. *SIGIR*, 2013.

[15] J. Teevan, K. Collins-Thompson, R. W. White, S. T. Dumais, and Y. Kim. Slow search: Information retrieval without time constraints. *HCIR*, 2013.

[16] E. M. Voorhees. Variations in relevance judgments and the measurement of retrieval effectiveness. *Information Processing & Management*, 36(5):697–716, 2000.

[17] J. Wang and D. Soergel. A user study of relevance judgments for e-discovery. *ASIST*, 2010.

[18] W. Webber and J. Pickens. Assessor disagreement and text classifier accuracy. *SIGIR*, 2013.