

## Powerful Geo-Search Using SRCH<sup>2</sup>

### Popularity of location-based services

We live in a data-driven world. Innovations in data storage, access, and retrieval have enabled an explosion of data-driven applications. New platforms, such as mobile and tablets, have put data on center-stage. Access to it, with a rich feature set including rapid geo-search, has become a must-have requirement.

If the first few decades of the Internet development have focused on amassing information and breaking the localities of information by making information accessible from anywhere, the next many decades will emphasize organizing information into geo-spaces to make information from anywhere "aware" of its localities. Many types of information including commercial information, consumer information, demographic information, and government information are only meaningful with their precise location defined. There is increasing popularity of location-based information and services, especially for mobile applications. More than 60% of the on-the-go mobile users are willing to share their location in order to receive more relevant content.

At the same time, it is becoming increasingly clear that providing intelligent access to location-based information is far more challenging than just providing access to information as usual. A key question remains to any information service provider: is your company able to satisfy your customers by making valuable information easily and intelligently accessible to them?

At SRCH<sup>2</sup>, we believe that existing search technologies, including the most widely-used open sourced solutions, fail to provide a satisfactory answer to this vital question. Such failures are costly to businesses because they lead to customers' dissatisfaction and loss of customers who eventually move away to competing services.

This white paper describes important features for today's search engines, introduces a new product that incorporates such features that can allow your users to search on your location-based information in an instant, error-tolerant way!

### Limitations of existing solutions in geo-search

Location-based search extends traditional keyword search by adding a very important location dimension. Since more and more users are searching information using mobile devices, a search engine needs to support this type of requests. In particular, mobile devices have a limited interface, making it much more difficult to type in keyword queries. In addition, often people do not know or remember the exact spelling of the search terms (such as business listings) they are looking for. Thus it is extremely important that the search engine support features that give users a pleasant search experience. Existing search solutions include using spatial search features of databases, and open-source search packages such as Lucene/Solr and Sphinx Search. These solutions may work for traditional keyword search, but cannot meet the new requirements by emerging applications.

It may be acceptable for a business to start with these ad-hoc search solutions when it is in the growing phase, since the solutions are readily available or can be quickly implemented. However, as the business flourishes, search quickly becomes a bottleneck in the growth and you have to look for better solutions.

It might be tempting to build in-house search solutions. But you may reach a different conclusion if you stop and think for a moment. What are your business goals? Where does your actual technical expertise lie? Are the efforts of building your own technology worth the time spent? Will your customers still be with you while you are busy improving their search experience, when you actually could have spent this valuable time working on your core business goals? More importantly, can you guarantee to deliver first-class search services after spending so much time and effort? These are latent costs that are not visible but impact your business a long way.

## The SRCH<sup>2</sup> solution

SRCH<sup>2</sup> knows search. SRCH<sup>2</sup> has invented and developed leading vertical search algorithms that offer unparalleled performance. We understand your search challenges and help you overcome them by allowing you to just focus on your business goals. We provide solutions to search your data the way your users want and give superior user experiences that are immediately recognized.

SRCH<sup>2</sup> supports powerful location-based search. Our patented technology enables incredibly fast full-text based instant search (a few milliseconds over tens of millions of records) on various kinds of data, including location-based information. Our product further supports error-tolerant, fuzzy search, which can find relevant answers even if there are errors in queries and data. These features are especially important for mobile users, where typing is time consuming and typos are more common. We have invested smartly in the world's smartest vertical search technology, so you don't have to invest heavily in hiring experts, training engineers, or building your own search infrastructure for your business to function efficiently.

SRCH<sup>2</sup> search is quick and easy, and does seamless integration to your system. The feature list includes:

*Instant search:* Also known as “search as you type,” “type-ahead search,” or “auto-completion,” instant search is a useful and popular feature for Web applications. SRCH<sup>2</sup> caches forward indexes in-memory, offering instant and relevant answers instantly as a user types in a keyword query character by character. It can provide quick feedback

that not only helps the users formulate a query quickly, but also helps them find answers easily.

*Fuzzy search:* SRCH<sup>2</sup> search also supports fuzzy search on your data. Because all forward indexes are cached in-memory, and SRCH<sup>2</sup> is able to traverse across indexed tries, it is able to identify misspelled businesses or individuals in ways that other technologies simply cannot. This feature is very important for a search engine, especially in mobile applications, since users are more prone to making typos during the search process due to the lack of a full-size keyboard and limited user input space. It is not uncommon to have geographic locations with complex spellings, making it difficult for users to find answers with traditional exact keyword search.

*Rapid geo-search:* SRCH<sup>2</sup> uses tree-based hybrid filtering to deliver faster and more relevant local search recommendations. With hybrid filtering, geo-locations are cross-indexed across indexed keyword nodes. Hybrid forward indexing is a new way to search your data, and doesn't depend on massive query logs (like Google's) to work. That means you get better local search recommendations than anyone else, right out of the chute.

## Feature comparison

SRCH<sup>2</sup> leads with significant margins over other solutions in the market both in terms of service and as a product. While the existing solutions are still struggling to introduce the basic forms of the above-mentioned features, SRCH<sup>2</sup> already has them available, with significant performance upgrades.

Table 1 compares our location-based features with Lucene/Solr, one of the most popular open-source solutions:

Feature	Lucene/Solr	SRCH <sup>2</sup>
Keyword search	Supported	Extremely fast
Instant search	Slow, poor ranking*	Extremely fast
Fuzzy search	Slow, poor ranking*	Extremely fast
Instant and fuzzy search	Slow, poor ranking*	Extremely fast

Table 1. Feature comparison of Lucene/Solr and SRCH<sup>2</sup>'s support of location-based search (\*: see the experimental results below)

We conducted experiments to compare SRCH<sup>2</sup> and Lucene/Solr for location-based instant and fuzzy search. We used a machine with an Intel Xeon E5620 2.4 GHz processor and 4GB of RAM, running a Ubuntu operating system. We used a data set with 5 million records of real business listings. We generated 500 queries with a single keyword and multiple keywords each. Every query had a 10km-by-10km spatial range. In Lucene/Solr, the location attribute representing the spatial condition was of a *LatLon* type field, and it used a *geofilt* spatial filter.

We used Lucene/Solr to support instant search by issuing wildcard queries (e.g., "ghira\*"). Our experiments on the data set ranging from 0.5 million to 5 million records showed that it is very inefficient, ironically making instant search very slow! Fuzzy search was even slower than instant search, making it impossible to be used in real applications.

Figures 1 and 2 compare the performance of SRCH<sup>2</sup> and Lucene/Solr for single-keyword instant searches and multiple-keyword instant searches, respectively. The x-axis represents the number of records, and the y-axis represents the average search time for queries. On an average, Lucene/Solr took 141 ms and 209 ms to perform a single-keyword search and multiple-keyword search on 5 million records, respectively. SRCH<sup>2</sup>'s solution took only 27.5 ms to 32 ms for each query.

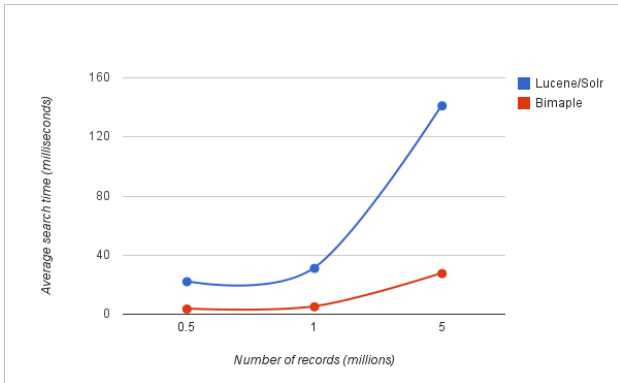


Figure 1. Single-keyword, instant search

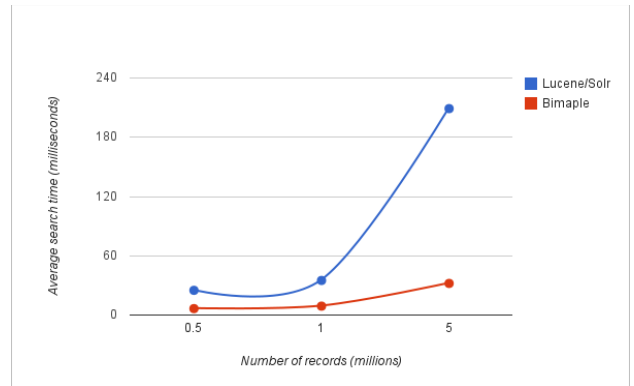


Figure 2. Multiple-keyword, instant search

Figures 3 and 4 show the performance comparison for fuzzy search. On an average, for keyword queries with up to 2 typos, Lucene/Solr took around 1,007ms (for single-keyword queries) and 1,040 ms for multiple-keyword queries. SRCH<sup>2</sup> returned results in only 12ms and 19 ms, respectively.

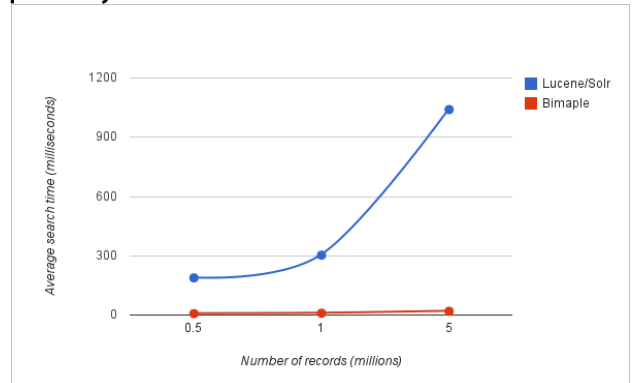


Figure 3. Single-keyword, fuzzy search

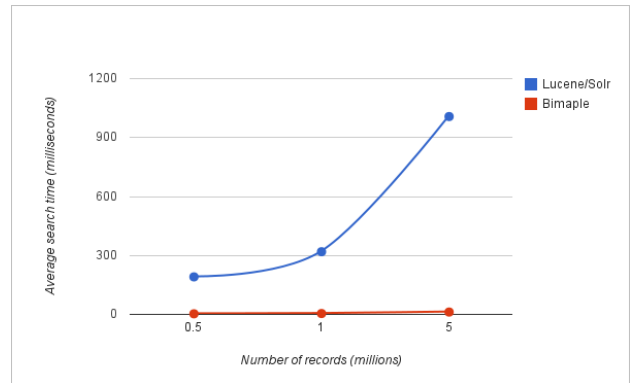


Figure 4. Multiple-keyword, fuzzy search

## Conclusion

As location-based services become increasingly more popular and important, applications need better search solutions to make their information more accessible, especially for mobile users. SRCH<sup>2</sup>'s location-based search solution meets these new requirements by enabling powerful features such as instant search and fuzzy search.



SRCH<sup>2</sup>: 2401 TechPortal, Calit2 Building, UC Irvine - Irvine, CA 92697 USA Tel +1-949-407-9971 <http://srch2.com>  
Copyright © 2012 SRCH<sup>2</sup>. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. SRCH<sup>2</sup> is a service mark of SRCH<sup>2</sup> Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.