Text Search with Spatial Constraints

Amber Han and Bradford G. Nickerson

Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada

Motivation

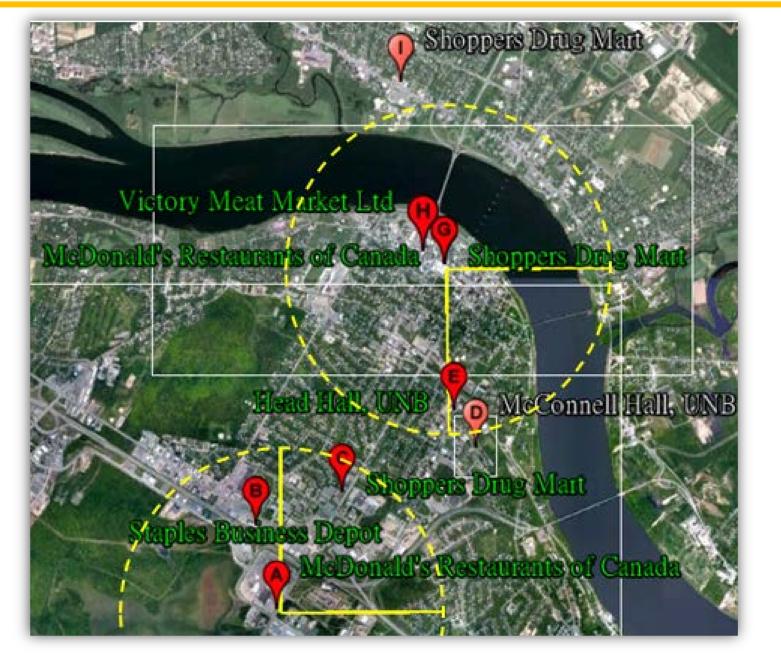
UNB

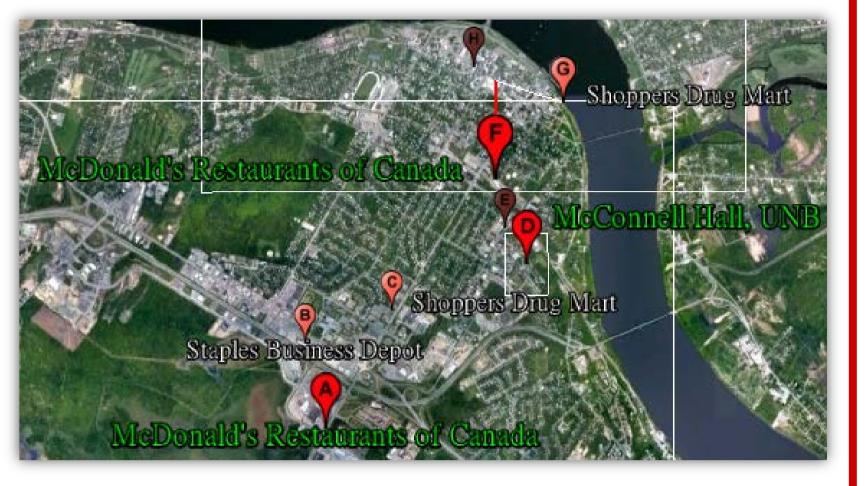
Improve search engines to better support I/O-efficient text + spatial queries.

Sample Data Record 1 Record 4		
Sample Data	Record 1	Record 4
	McDonald's Restaurants of	Shoppers Drug Mart
	Canada	(45.942639N, 66.655147W) C
Records	(45.934516N, 66.663308W) A	(45.961817N, 66.643622W) G
consist of	(45.961817N, 66.643622W) F	
plain text and	Record 2	(45.976414N, 66.649003W) I
associated		Record 5
spatial	McConnell Hall, University of New Brunswick	Staple Business Depot
information	(45.946419N, 66.639297W) D	(45.939778N, 66.662633W) B
in a database	Record 3	People R
	Llood Lloll Lloix araity of Novy	Record 6
	Head Hall, University of New Brunswick	Victory Meat Market Ltd
	(45.949961N, 66.641711W) E	(45.962739N, 66.645572W) H

Three query types

Q1 = (t): search returns a ranked list of items matching the search string t, along with their associated spatial information (e.g. latitude, longitude).
Example query: Q = ("Mc") Records A, F, D are returned.

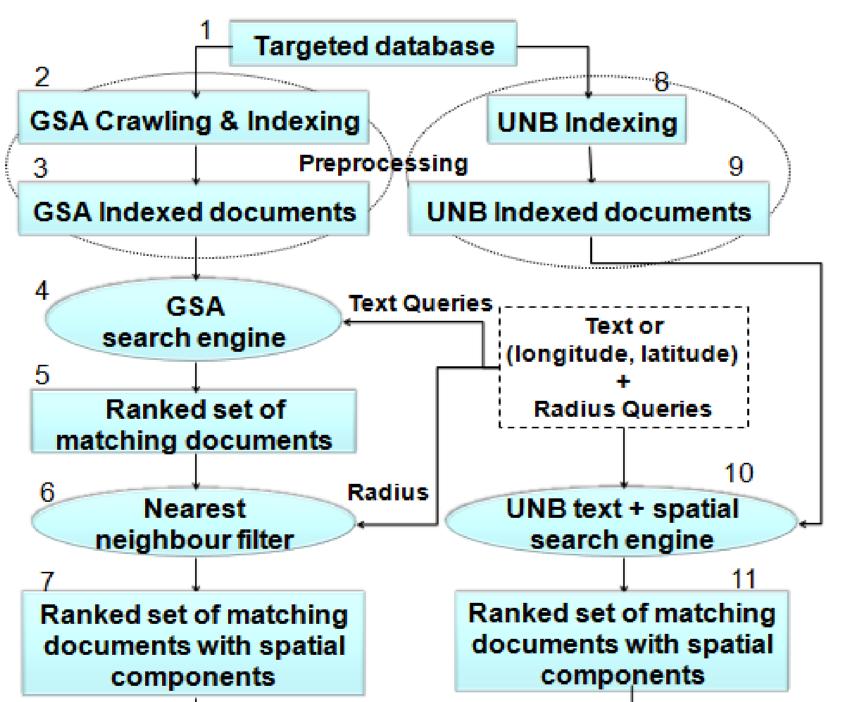




Q2 = (t, r): search returns a

Test Methodology

The system will be tested on the UNB Connell Memorial Herbarium database (N = 40,791). The performance of the system will be compared with the Google Search Appliance (GSA) on the targeted data set.



ranked list of documents with at least one spatial component having its location falling within the circle of radius r centred at the position p of the ranked documents matching t.
Example query: Q2 = ("McDonald", 1.5km)

Q2 = ("McDonald", 1.5km) Records A, F, G, H, B, C, E are returned.

Q3 = (p, r): search returns a ranked list of documents with at least one spatial component having its location falling within the circle of radius r centred at position p.

 Example query: Q3 = (position P = (45.952567N, 66.646001W), 1.5km)
 Records E, D, F, G, H, C are returned. Victory Mean Minace Lol Shippers Drog Minit Position (15.957/507/3, 66.64/600179) Elead Hall, UNB Shoppers Drog Man Shoppers Drog Man

In Q2 and Q3 search results, the points nearer to the point(s) of interest are ranked higher



Example record from the UNB Connell Memorial Herbarium database with the query string "Asteraceae" 12 Compare and analyze

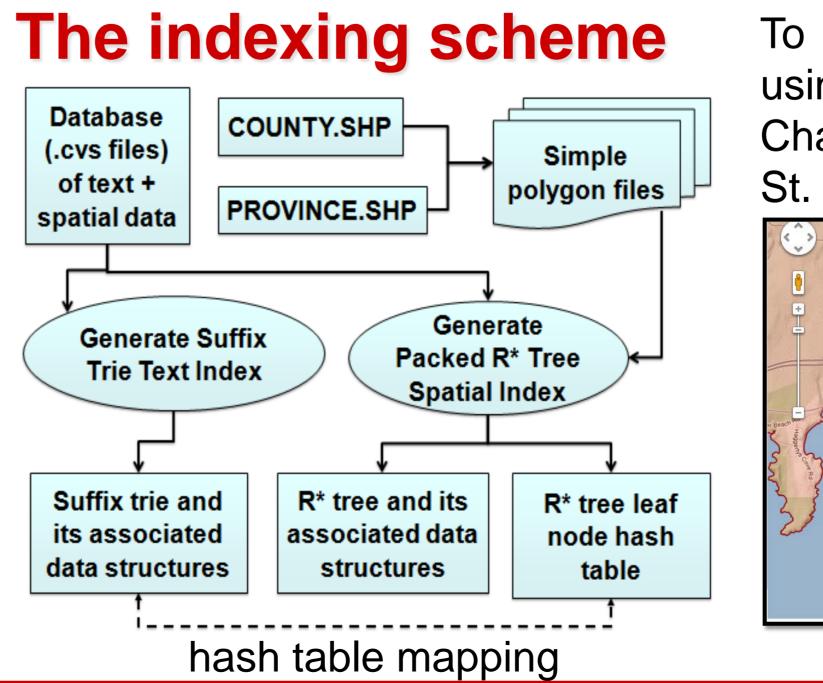
Example result from the GSA on the UNB Herbarium data set with query string "Asteraceae"



UNB Accession Number: 18395 Basic Information, FullName, Matricaria discoidea DC. Family, **Asteraceae**. Synonyms, Matricaria matricarioides auct. non (Less.) Porter; Chamomilla suaveolens. ... 131.202.243.11/test/HerbariumDatabase/18395.htm - 2012-12-31 - <u>Cached</u>

Record 14634

Specimen Label Data for the Connell Memorial Herbarium. UNB Accession Number: 14634 Basic Information, FullName, Sonchus arvensis L. Family, Asteraceae. Synonyms, French Name, Laiteron des champs. ... 131.202.243.11/test/HerbariumDatabase/14634.htm - 2012-12-31 - <u>Cached</u>



To reduce space, we simplify the polygons using the **PackPolygon** algorithm. Charlotte (left) $27,756 \rightarrow 500$ points St. John (right) $25,774 \rightarrow 500$ points

