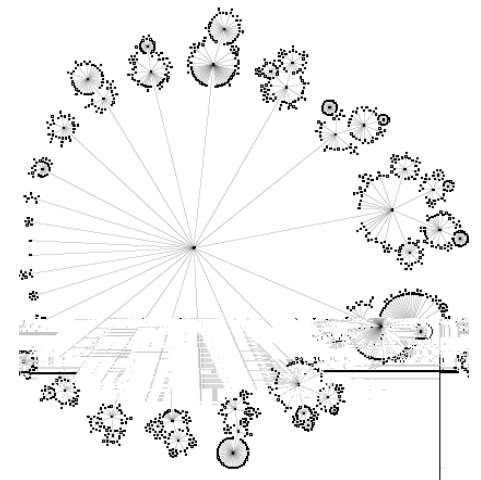


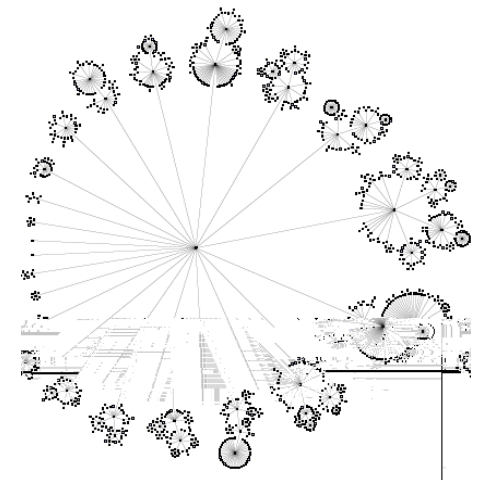
Location semantics in web-multimedia content

Harish Katti



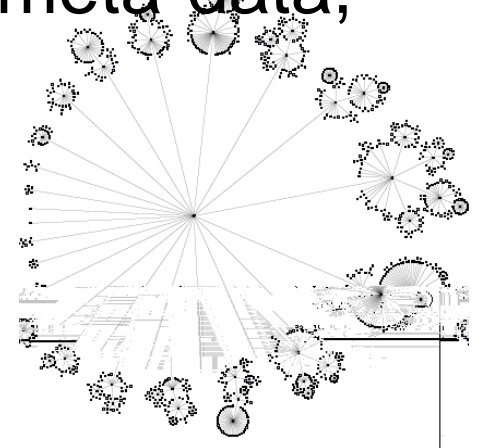
The problem

- **Given** a multimedia content about places and their salient features (image, text, metadata)
- **To find** out whether a query (multimedia, web page content) talks about a particular place and if so, which aspects ?
- Project funded by the Singapore Publishing Group



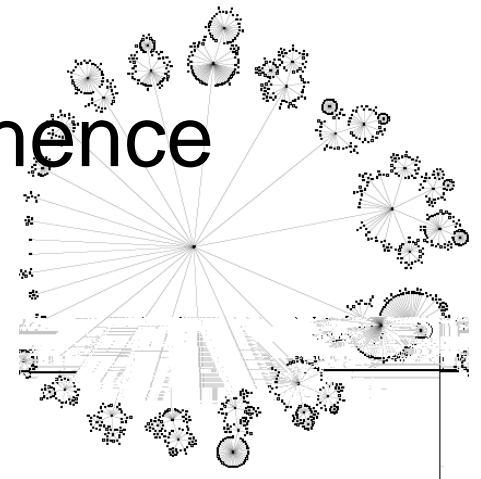
How is it different from usual search ?

- Content with many implicit assumptions: Blogs, casual write-ups. Eg; wikipedia Vs travel-blog
- Cannot rely only on information like domain names and IP addresses as content maybe unrelated (eg; citeseer mirror hosted at NUS)
- Does not make use of image content, meta-data, etc.
- Usually doesn't rely on semantics



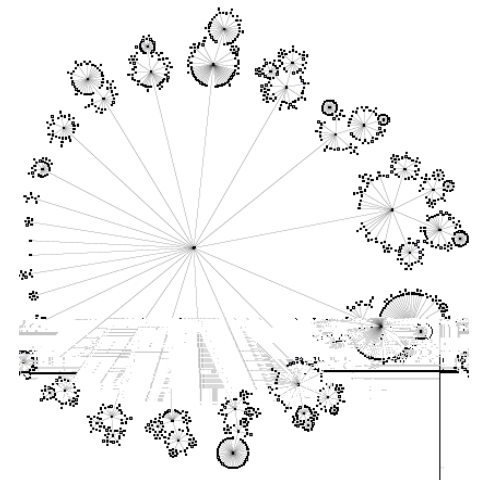
Why do this ?

- Interesting: Multimedia search and involves semantics
- Difficult:
 - Scale of the problem
 - Heterogeneous data
 - Non-uniform information distribution
- Useful: People are increasingly comfortable with multimedia and hence content explosion.



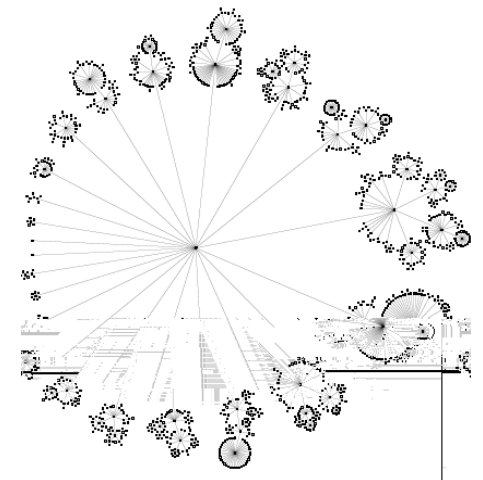
Our approach

- Multimedia representation as a directed, semantic graph and finding the best matching context in the representation for the query (graph)
- Identifying salient features for the nodes from multimedia content is a separate problem



Node information

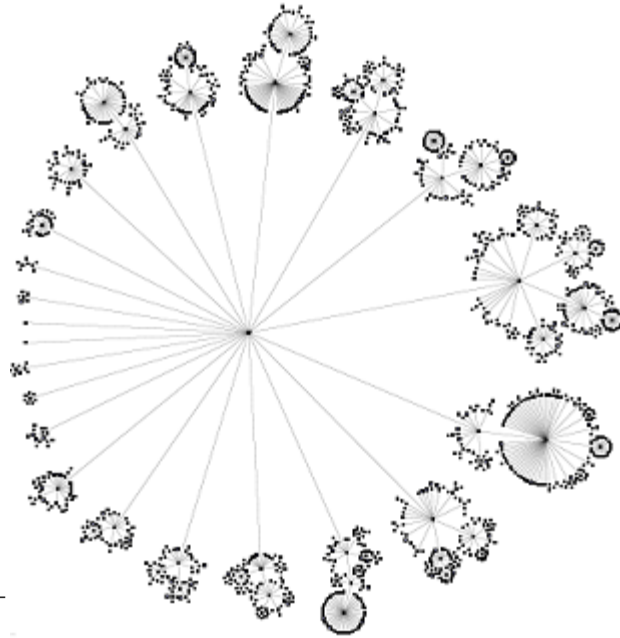
- Phrases, words. eg: “Marina Bay”.
- Metadata like (longitude, latitude), any other geo-spatial terms. EXIF information.
- Images relevant to that concept from flickr
- Salient features for this node, some of these features may also have links to other



4 level sub-category taxonomy of “Singapore” in wikipedia

Level 1

Level 2



Singapore'

"Buildings_and_structures_in_Singapore"

"Economy_of_Singapore"

"Education_in_Singapore"

"Environment_of_Singapore"

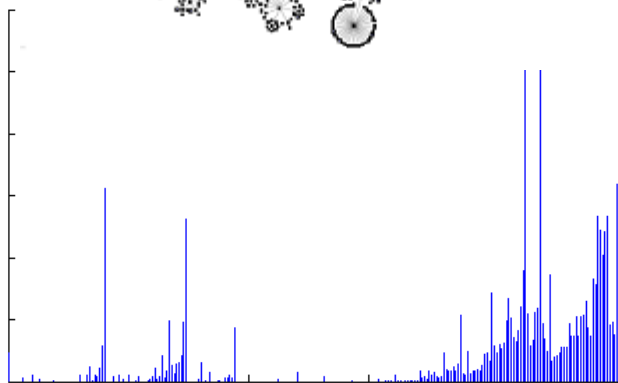
"Geography_of_Singapore"

"Government_of_Singapore"

"Health_in_Singapore"

"History_of_Singapore"

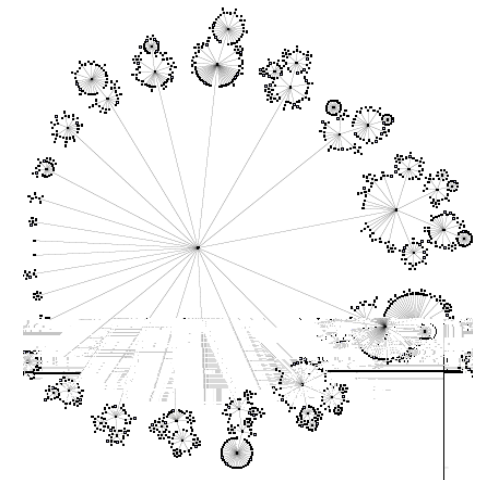
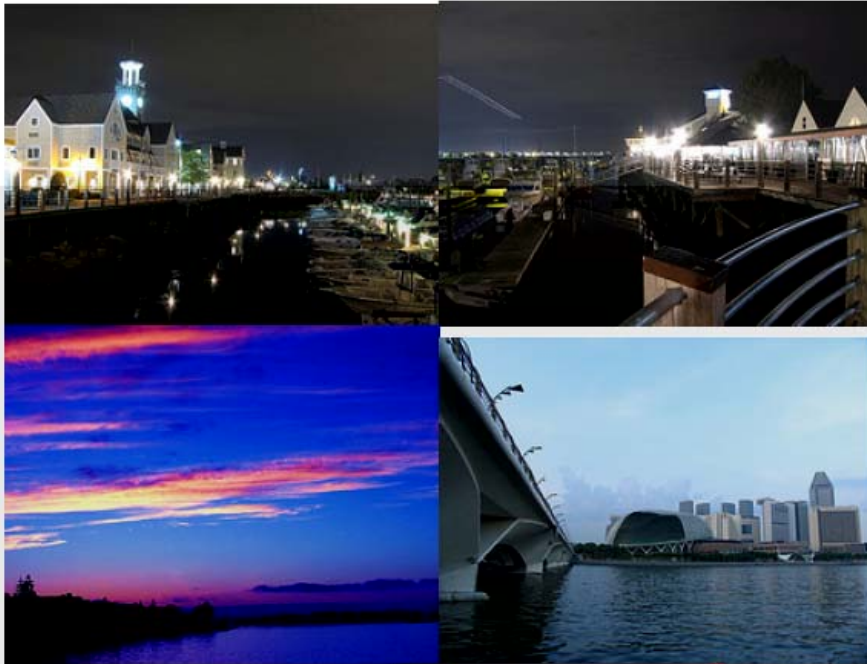
"Images_of_Singapore"



Node frequency histogram
(information distribution)

Example: Marina Bay

- http://en.wikipedia.org/wiki/Marina_Bay%2C_Singapore
- The categorical context could come from the wiki taxonomy:
 - Categories: Downtown Core | Marina Bay | Places in Singapore | Bays



- Ethical issues with content
- Reliability of information
- Quality of information
- Good problem
- Automatic updating of the representation so that it stays up to date
- Automatic evaluation of the memory-snapshot.. Self healing, self learning...