

Sample of Level 3

Field of research: Linguistics and Information Management

Geographic Information Retrieval and Text Mining on Chinese Tourism Webpages

1. Introduction

The ~~state of the~~ Internet ~~is characterized~~ has by an abundance of ~~information and data that has and~~ become one of the most significant resources ~~of message for the majority of people in their in~~ our daily lives. ~~Such. These~~ resources contain an enormous amount of ~~facts or data, the retrieval of~~ which depends mostly on ~~the~~ webpage search engines ~~in order to effectively retrieve and search for the find a suitable answer to the an inquiry~~. However, most of this knowledge comprises ~~of not only either~~ non-structured ~~data or but~~ also semi-structured data (Mitra *et al.* 2003), ~~but also and~~ at ~~the~~ precurrent ~~moment,~~ the ability of regular search engines is only limited to the retrieval of ~~a the~~ basic keywords, ~~instead of the rather~~ than analysis of ~~a the~~ subject matter and content ~~and topic~~ of the webpage itself; this application then, is ~~yet~~ still far from perfection. ~~Owing to For~~ the reasons mentioned above, ~~there are many much~~ researches on efficient message and data extraction ~~has~~ ve been ~~applied conducted focussed on into to the topic on the effective~~ management ing these of facts and data ~~effectively while associated with efficient carrying out the message and data extraction efficiently.~~

~~From these This~~ research has resulted in es, ~~there are~~ many developments in information retrieval and data mining ~~strategie kills.~~ However, these ~~strategie kills~~ are mostly aimed ing at ~~the~~ semantic data only. Based on a recent estimation, ~~there are~~ about twenty percent of web-users ~~browse in their e~~ inquire for ~~based on~~ spatial context, such as search ~~looking for~~ restaurants, theatres or ~~,~~ academic institutions; in addition, eighty percent of ~~those~~ web-users ~~are typing~~ their queries in services with a more location orientation ~~based services~~ (Kornai 2001; Souza, 2005).

Comment [SM1]: CHECK: Some people make a clear distinction between these two terms, which might be very relevant to your theme. Considering furnishing a contrasting definition of the two terms unless they are already very clearly used and recognized within your discipline.

Comment [SM2]: CHECK: Again, what is your distinction between “facts” and “data”. It would be an option simply to leave out “factors or”...

such as ~~search~~ looking for New York delicacies or hostels in Portland, ~~and so on, etc.~~ However, they are hindered not only by the current ~~state of~~ development of the search engine's ~~ability~~, but also by the limitations ~~imposed by~~ ~~on the~~ differences between written English and Chinese. ~~Because~~, ~~considering that~~ it is not possible to put a space between written Chinese ~~characters, this~~, has ~~easily~~ caused ~~a~~ perplexities ~~in relation to~~ ~~y within the~~ word segmentation. ~~Therefore~~, ~~As for the~~ regular search engines ~~therefore this~~, it is quite a barrier to effectively indexing ~~the~~ web-content ~~written in the~~ Chinese ~~written language~~. For example, if the user is browsing for “Portland 民宿” [read as ‘min su’], ~~which this~~ means hostels in Portland; the regular search engine might only perform some search on either Portland or ~~on~~ 民宿 ~~separately from their~~ database. ~~This~~, ~~however this~~ inquiry ~~is therefore~~ lacking thematic and spatial context, ~~and this~~ ~~which~~ creates a big gap between the demand and supply of the query (Buyukkokten 1999).

Although ~~the~~ ~~best~~ Geographic Information System (GIS) has ~~the~~ ~~the~~ capability ~~of to~~ handle ~~ing the~~ geographic ~~data~~, ~~however the~~ access ~~to of the~~ spatial data is mostly limited to ~~spatial~~ coordinates created by ~~the~~ ~~geometricy~~ space expressions. Generally, people express their knowledge ~~in of~~ geographic locations ~~s~~ by using spatial content ~~xt~~ such as place names, labels, addresses or even telephone numbers, instead of using geometric coordinates ~~as it is being used in the GIS~~ (Jones 2001). The content of a webpage is a concrete example of this phenomenon, where people express ~~their location data in a spatial context.~~ ~~Band~~ by using this ~~implicit spatial data~~, the connection between webpages and any particular location can be established. Even though ~~Web GIS is presently available, at the moment, the present of Web GIS is there, however~~ it is quite difficult to combine ~~GIS~~ the analysis ~~ability of GIS~~ and text analysis, since ~~its the former's~~ usability does not surpass ~~the~~ traditional GIS.

Comment [SM3]: CHECK: While this is grammatical, it is not especially clear. I suggest you enlarge on what you mean here.

Comment [SM4]: CHECK: Again, why is the data “implicit”? explain this

Comment [SM5]: CHECK: I have assumed here that you mean Web GIS doesn't surpass the regular form, but it could have been read to mean that text analysis doesn't