

What?

Query segmentation refers to identification of meaningful units in a query. For example, "australian open" "home page". Query segmentation can help improve IR performance, query expansion and query suggestion. Past work has used a variety of resources such as content of Web documents, POS tagger and chunker, named entity lists, and human annotated data.

Why only Query Logs?

Queries have their own unique structure. Use of document text or human annotation for query segmentation increases the risk of projection of natural language structure onto queries. So the use of query logs as the only resource is the best and scalable approach to unravel units specific to queries.



Query Segmentation using only Query Logs

Nikita Mishra, Rishiraj Saha Roy, Niloy Ganguly (IIT Kharagpur)
Srivatsan Laxman and Monojit Choudhury (Microsoft Research India)

contact:
monojitc@microsoft.com
niloy@cse.iitkgp.ernet.in

1

X: observed counts of candidates.

leonardo da vinci: 5
xp vista: 2

2

Expected probability of candidate:

$$P_i = \frac{(\ell_i - n + 1) \times (\ell_i - n)!}{\ell_i!}$$

$$P(\text{leonardo da vinci}) = 1/30$$

$$P(\text{xp vista}) = 1/6$$

3

Expected counts

E(X):

leonardo da vinci: 7/30

xp vista: 7/6

4

Prune the candidates by using *Hoeffding's inequality*:

$$\text{Prob}[X \geq N] \leq \exp\left(-\frac{2(N - E(X))^2}{k}\right)$$

$$p[\text{leonardo} \dots \geq 5] = 0.015$$

$$p[\text{xp vista} \geq 2] = 0.82$$

xp vista → not a segment

5

Given a query, segment it such that the sum of the negative log-probabilities (obtained in step 4) of all the segments is maximized.

leonardo da vinci artwork

houses for sale in things to do in difference between information about history of the how to make a list of jobs in spot a fake pictures of how to use what is the very hungry caterpillar gold coast nikon D5000 windows media player

1 gram equals how many pounds
16 sherwood crescent narraweena richard
1991 mazda 121 sevice manual download
gui differences between win xp and vista
2008 chrysler town and country lx
2012 in 3d at hoyts norwood
abraxane and indolent breast cancer resp
complete illustrations of leonardo da vinci
accounting information system pearson p
activities for kids 4 years old
adelaide elder park new year eve
halo 2 memory requirements in xp vista
advanced diploma in recreation studies
leonardo da vinci life and times
after school because of you lyric
alaska mental health trust land office
leonardo da vinci birthplace
all american food buffet marietta ga
amf ten pin bowling results sa
security issues in windows vista and xp
goes here Examples of queries goes here
Examples of queries goes here Examples
leonardo paintings seen in da vinci code
another name for the kite shape
apartments for rent in boone iowa
cost comparisons xp vista win 7
are hosta plants hard to grow
arrivals charles de gaulle airport paris
astrology horoscope year of the ox
machines designed by leonardo da vinci
attract positive people into your life
australia news at the winter olympics
bsod on boot up both xp and vista
australian governments response to david
auth ugg black kensington leather boots
average temperature in july in countries
complete biography of leonardo da vinci
baba khan ft fatman scoop tonight
backpackers guide san juan puerto rico
ubuntu dual boot with win xp and vista
ban than chali dekho mp3 download
is leonardo the real name of da vinci
barn and line dance caller centre
battlefield bad company 2 demo torrent
beat the sensei in club penguin
error in xp and vista importing cell pics
teachers employment in nsw
of the australian gold rush
act digital waterproff camera
last night of the proms

