

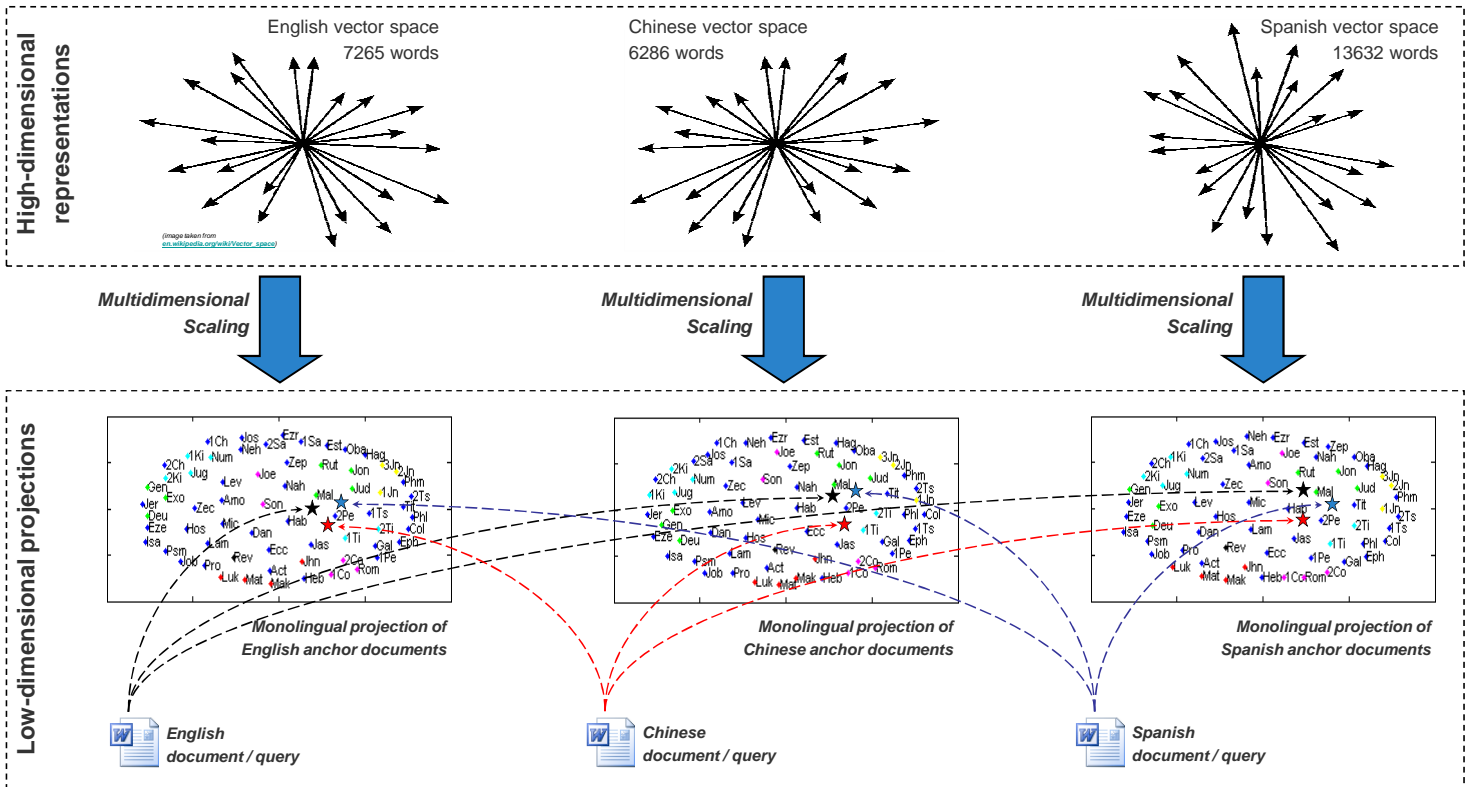


Exploiting Multidimensional Scaling Projections for Cross-Language Information Retrieval

Rafael E. Banchs & Andreas Kaltenbrunner

Barcelona Media Innovation Centre, Ocata #1, 08003 Barcelona, Spain. (rafael.banchs/andreas.kaltenbrunner@barcelonamedia.org)

Trilingual parallel collection: English, Chinese, Spanish
(66 books from The Holy Bible)



Probe document / query placement method

1.- Projection matrix computation:

$$T = M D^{-1}$$

M : coordinates of anchor documents in the projected vector space
 D : distances among anchor documents in the original vector space

2.- Probe document / query projection:

$$m = T d$$

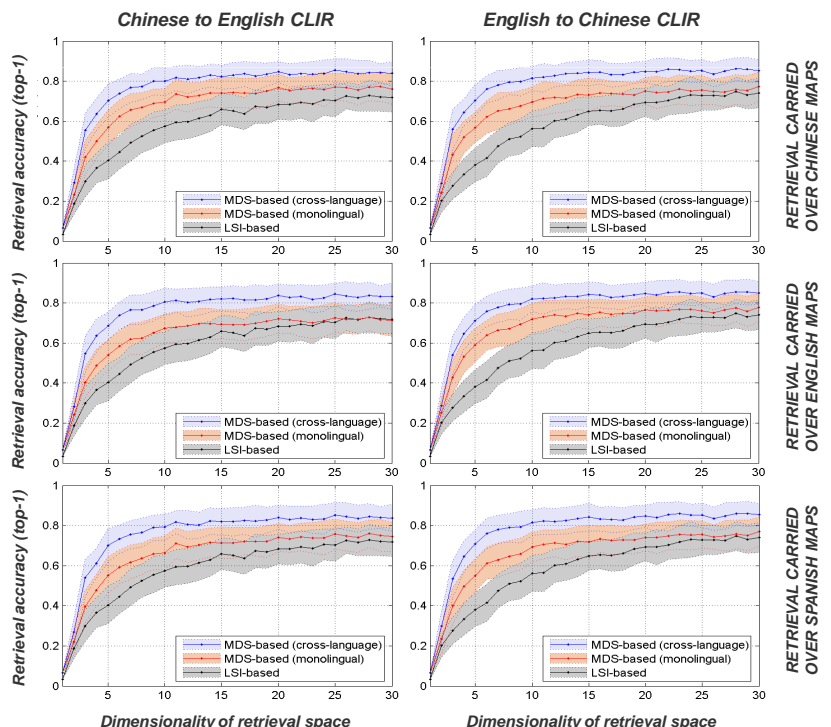
d : distances between probe document / query and anchor documents in the original vector space of source language
 m : coordinates of probe document / query in the projected vector space of retrieval language

3.- Two variants of the method:

- 3.a.- Monolingual projection matrix:
 M and D computed on retrieval language
- 3.b.- Cross-language projection matrix:
 M computed on retrieval language and D computed on source language

Proposed method compared with Latent Semantic Indexing...

(30 anchor documents / 30 probe documents / 100 random selections)



to be continued...