

[usofan@kashanu.ac.ir](mailto:usofan@kashanu.ac.ir)

---

- stop words

( )

... this by the

[ ]

[ ]

TREC

... « » « » « » « »

[ ] zipf

frequency . rank  $\approx$  constant

( )

[ ]

[ ]

zipf

[ ]

[ ]

« »  
: ... « » « » « » « » « »

...

« » « » (ZWNJ)  
 « »  
 ( )  
 ( )  
 [ ] « »

[ ]

)

(

high jump ape\_man bathroom  
 community center finance committe

[ ]

» feuerundlebensversicherungsgesellschaft «      » Versicherungsgesellschaft  
 [ ] «

« »      « »      « »

)      « »      « » (

« »	←	« »+	« »
« »	←	« »+	« »
« »	←	« »+	« »
« »	←	« »+	« »
« »	←	« »+	« »
« »	←	« »+	« »

utf8



[ ]

( )

(if else )

Ø/ \_ / « »/ + « »

windows unipad linux mandrake 9.1 gedit

« \_ \_ » (-)

(«»)

[ ] : Ø

( )

/\* \*/

( )

(or )/

[ ]

(=)



( ) (+)

( )

« »/« »/« »/« »/« »/« » = \_ \_  
« »/« »/« » = \_

∅ / « » + « » / « » / « » = \_ \_  
∅ / \_ \_ / \_ \_ / \_ \_ + « » = \_ \_

« »

:  
∅ / \_ \_ + « » + / « »  
... « » « »

( ) htrack ( )  
( ) (Windows 1256 ) CP 1256

utf-8


( )

( )

\_ salton , G. M.J. McGill(1983). An introduction to Information retrieval. McGraw Hill.

\_ I.witten , A.Moffat and T.Bell. Managing Gigaytes : Compression and indexing documents and images. margon kaufmann , 2nd edition 1999.

\_ Luhn , H.P.(1957) . "A statistical approach to mechanized information encoding and searching of literary information". IBM journal of research and development , vol .1 , no 4, page 309.

" ( )

«http://www.itrc.ac.ir»

( ) ( ) ( )

: ( ) ( )

( ) :

: ( )

( )

: ( )