

ABSTRACT

This research is one of the important steps towards processing the most significant challenge in digital libraries and web, which is computing document's rank, its importance, and its relevance to the user information need. This is achieved through the utilization of fuzzy logic high potential capabilities in dealing with such sort of problems and providing notable flexibility for the user to get his favorite subjects.

The research is concerned with designing and implementing a proposed information retrieval system, called FIRS (Fuzzy Information Retrieval System). This system is developed to deal with huge database, which contains different text file types and sizes. This database are distributed over a collection of server computers, connected with the intranet network that are dedicated for this system.

The system has the ability for mining of data available in this database and retrieving the useful information that corresponds with the user need. This is accomplished through applying the proposed algorithms for indexing process, document's rank computations, generating keywords process, and finally, displaying information retrieval results. The proposed system gives high quality results comparing with other information retrieving algorithms.

(Fuzzy Information Retrieval System) FIRS

(servers)

"Surveys"

() (Lark Survey)

[2].())

()

(PC)

:

.

.

.

.

.

Circulation System

Punched Card

Digital Libraries

Melvin J. Voight .

Clay L. Perry .

()

(Integrated Systems)

National Library of Medicine

Medical Literatur Analysis and Medlars Retrival System

Meoicus Index

SDI

Nasa

[2].

:

:

:

[10].

()

()

)

[6]"

.("

[4]

:

[5]()

[9] :

[7] Zadeh (Soft Computing-SC)

SC

SC

(rough sets)

[8]Bayesian

L. A. Zadeh

(computer vision)

[3]

[8].

(E.Book)

: [1]

" "

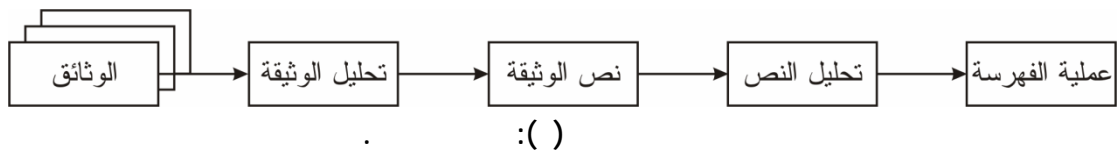
(Hard)

Fuzzy Information Retrieval System (FIRS)
()

(:)
()

()

()



:

()

معلومات خاصة بالكلمة	معلومات خاصة بالوثيقة
الكلمة	اسم الوثيقة
عدد تكرارها الكلي	امتداد الوثيقة
الوثائق التي تظهر فيها	مسار الوثيقة
عدد الوثائق	عنوان الوثيقة
اكبر تكرار للكلمة	URL
ابعد مكان تظهر فيه	عدد URLs
	الكلمة في العنوان
	الكلمة في المحتوى
	اول ظهور للكلمة
	حجم الوثيقة
	تسلسل الكلمة في الوثيقة
	تكرار الكلمة في الوثيقة
	MF_count_low
	MF_count_high
	MF_loc_far
	MF_loc_near
	MF_URL_low
	MF_URL_high
	RANK

:()

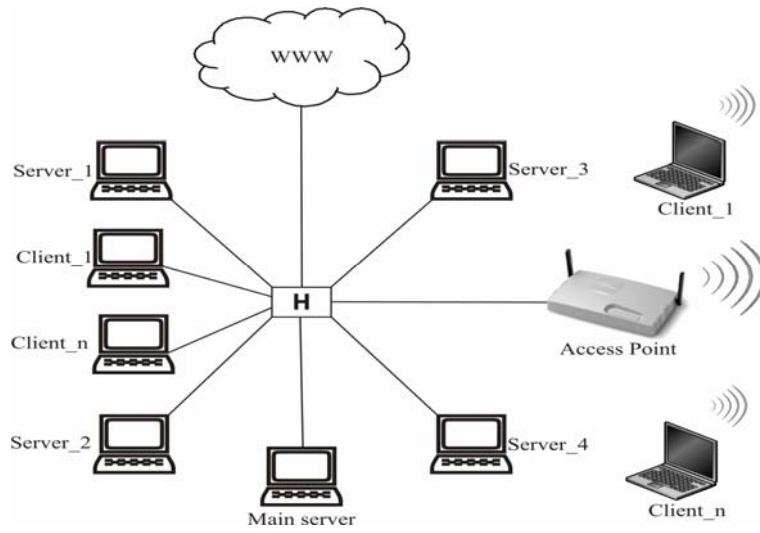
()

(Main Server)

(Clients)

HUB

Access Point



.FIRS ()

Interfaces

(clients) FIRS (servers)

HTML

tags

.regular expression

special characters

()

Boyer-

.Moore

:Fuzzification



()

$MF_count_low = 1 - (1 / (1 + \exp(-((count - (max_count / 2))))))$ 1

$MF_count_high = 1 / (1 + \exp(-((count - (max_count / 2)))))$ 2

$MF_loc_near = 1 - (loc / end_loc)$ 3

$MF_loc_far = loc / end_loc$ 4

$MF_URL_low = 1 - (1 / (1 + \exp(-((fileURLcount - (max_no_of_URLs / 2))))))$. .. 5

$MF_URL_high = 1 / (1 + \exp(-((fileURLcount - (max_no_of_URLs / 2)))))$ 6

:MF_count_low

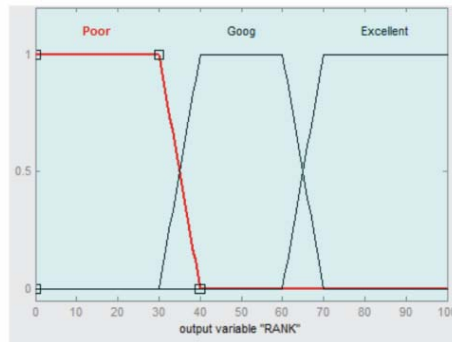
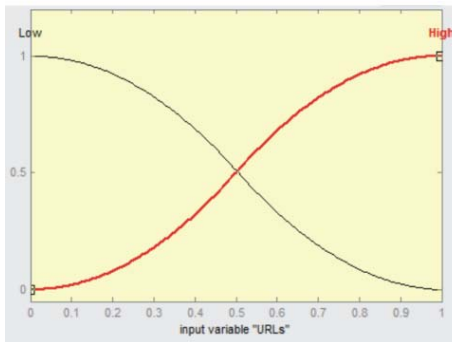
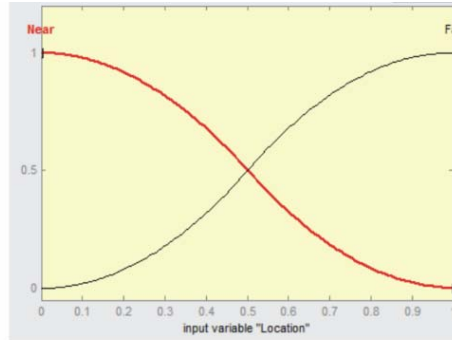
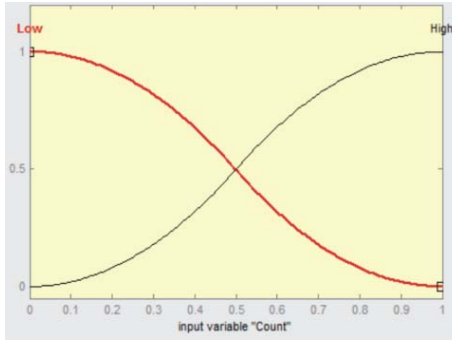
:MF_count_high

:MF_loc_near

:MF_loc_far

:MF_URL_low

:MF_URL_high



:()

:Knowledgebase Rules



(poor, good, excellent)

HTML

.()

1. If (Count is Low) and (Location is Near) then (RANK is Goog) (1)
2. If (Count is Low) or (Location is Far) then (RANK is Poor) (1)
3. If (Count is High) and (Location is Near) then (RANK is Excellent) (1)
4. If (Count is High) and (Location is Far) then (RANK is Goog) (1)

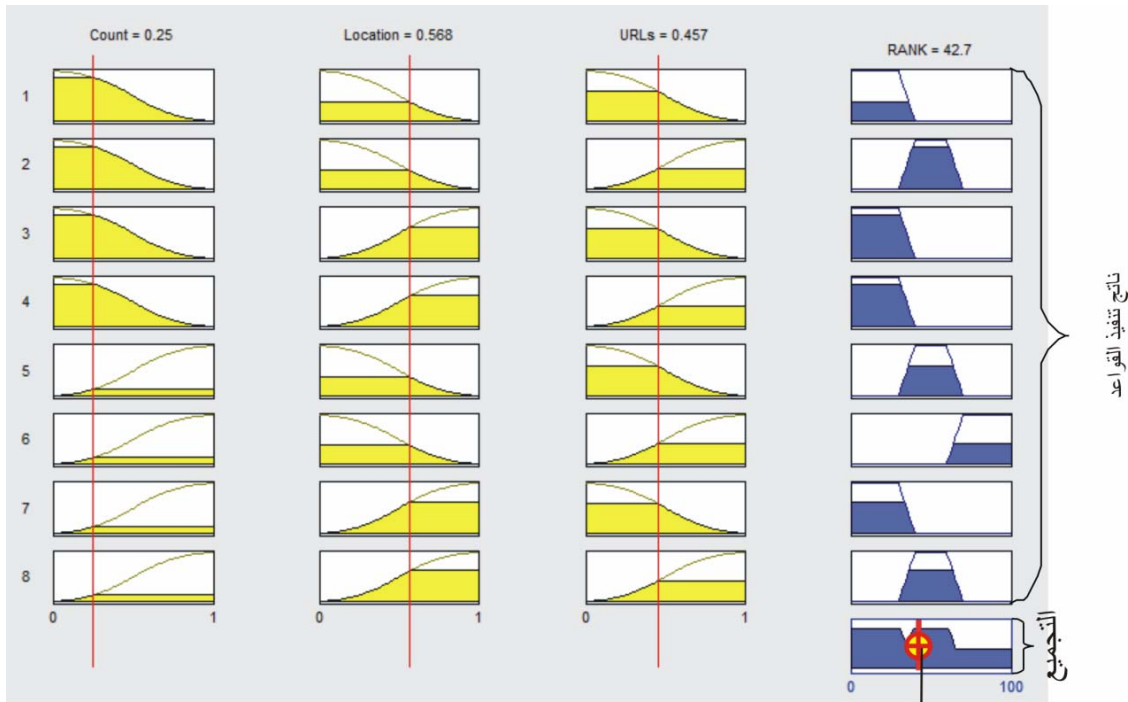
1. If (Count is Low) and (Location is Near) and (URLs is Low) then (RANK is Poor) (1)
2. If (Count is Low) or (Location is Near) or (URLs is High) then (RANK is Goog) (1)
3. If (Count is Low) or (Location is Far) or (URLs is Low) then (RANK is Poor) (1)
4. If (Count is Low) or (Location is Far) or (URLs is High) then (RANK is Poor) (1)
5. If (Count is High) or (Location is Near) or (URLs is Low) then (RANK is Goog) (1)
6. If (Count is High) or (Location is Near) or (URLs is High) then (RANK is Excellent) (1)
7. If (Count is High) or (Location is Far) or (URLs is Low) then (RANK is Poor) (1)
8. If (Count is High) or (Location is Far) or (URLs is High) then (RANK is Goog) (1)

() :

Aggregation ❖

() :

OR-MAX



حساب قيمة الرتبة باستخدام طريقة centriod () :

:Defuzzification ❖

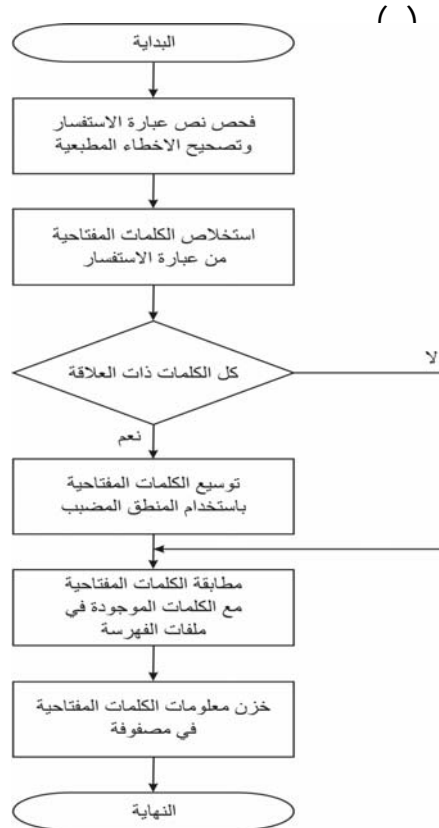
Rank

: () centriod

$$RANK = \frac{\sum Xi \mu(X)}{\sum \mu(X)} \dots\dots\dots \gamma$$

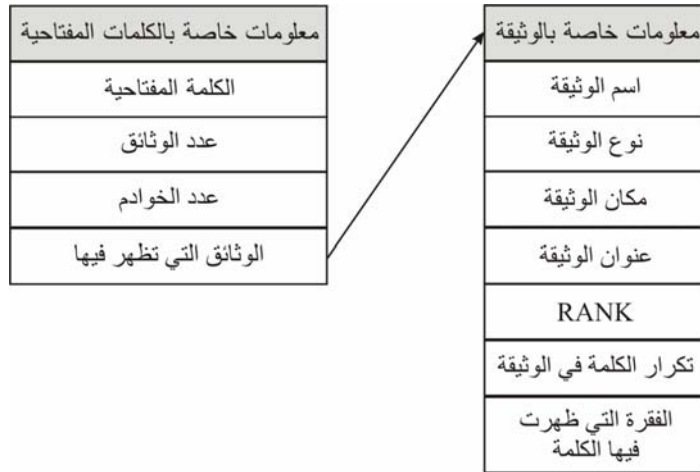
Search

SmartTextBox



: ()

() .



() :

tags

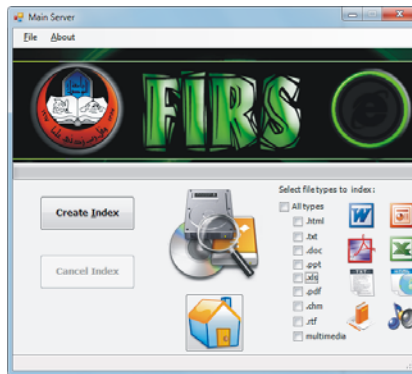
Allrelevant

CheckBox

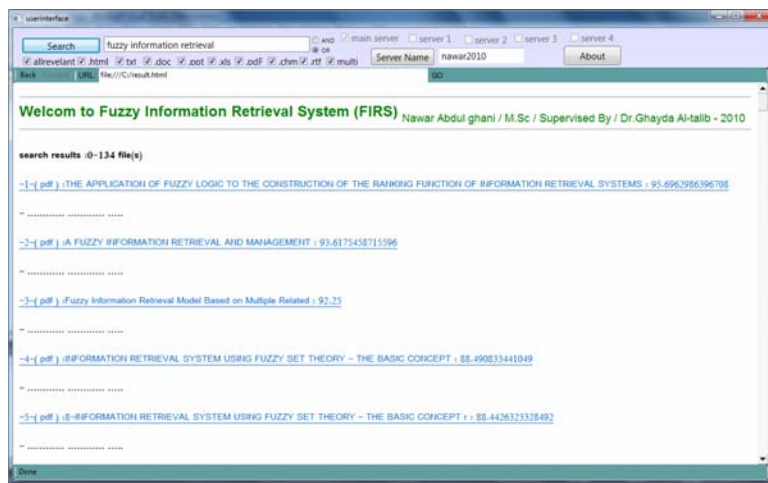
Levenshtein-Distance

Levenshtein-Distance

	FIRS	.
		FIRS
		.
Visual Studio 2008	C#	
	.NET Framework Ver. 3.5	
.windows sockets		
multithreading		C#
\		
	:	
		-
64bit	8GB	RAM
.320GB	2.9GHz	
		-
	.NET Framework Ver. 3.5	-
	FIRS	-
		.
wireless access point		-
		-
()		
		.
()		()



:()



:()

C#

lucene

.()

()

()

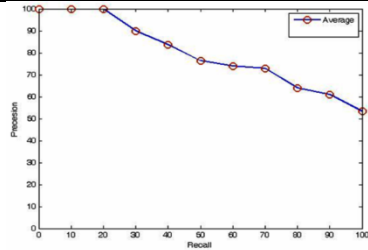
()

.()

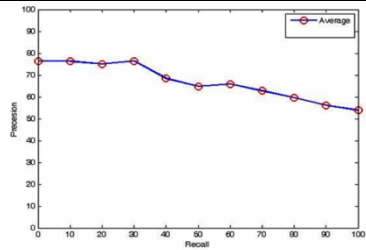
Lucene

:()

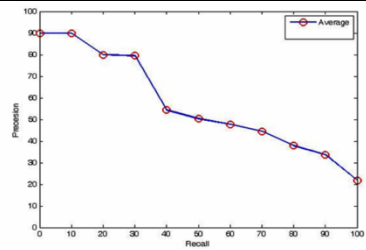
Recall	Precision	Precision Lucene	Precision
10%	76.25%	90%	100%
20%	75.13%	79.98%	100%
30%	76.4%	79.5%	90%
40%	68.4%	54.37%	83.7%
50%	64.8%	50.38%	76.4%
60%	65.9%	47.82%	73.9%
70%	62.7%	44.57%	72.9%
80%	59.6%	37.96%	64.03%
90%	56.2%	33.73%	61.1%
100%	53.8%	21.82%	53.4%



أ



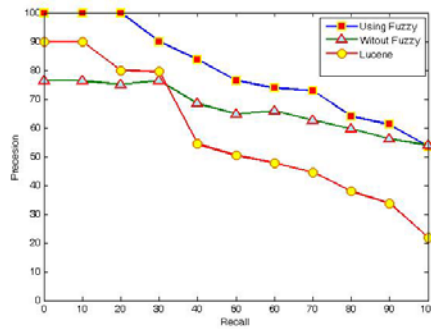
ب



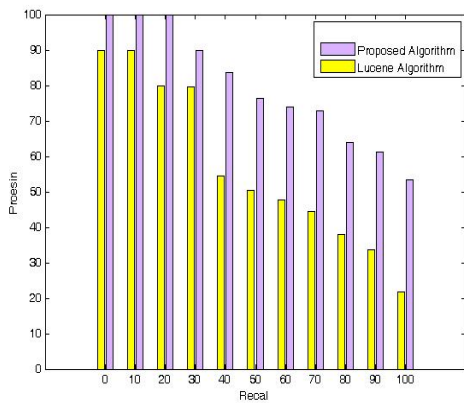
ج

:()

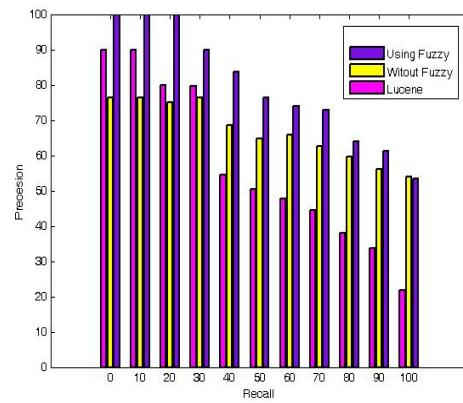
.Lucene



:()



:()



:()

.Lucene

(Regular Expression)

Boyer-Moore

Levenshtein-Distance

" " []

.

[٢]

-

- [3] Acharya, T. and Ray, A. K., 2005, "Image Processing: Principles and Application", John Wiley & Sons, Inc, Hoboken, New Jersey.
- [4] Bhaskar Karn , 1998 ,” Information Retrieval System Using Fuzzy Set Theory – The Basic Concept”, Assistant Professor, Department Of MIS(Management Information Systems), Birla Institute Of Technology, Mesra, Ranchi.
- [5] D. Hiemstra , 2001, "Using Language Models for Information Retrieval", Universiteit Twente. ISBN 90-75296-05-3, ISSN 1381-3617; No. 01-32 (CTIT Ph.D. Thesis Series) Subject headings: information retrieval, natural language processing.
- [6] Fabio Crestani and Gabriella Pasi, 1997, "Soft Information Retrieval: Applications of Fuzzy Set Theory and Neural Networks" , Department of Computing Science, University of Glasgow.
- [7] L.A. Zadeh, 1994, “Fuzzy logic, neural networks and soft computing”, Communication of ACM.
- [8] Ricardo Baeza-Yates, 2003, " Information retrieval in the Web: beyond current search engines " , Center for Web Research, Department of Computer Science, University of Chile, Blanco Encalada, 2120 Santiago, Chile , Received 1 January 2003; accepted 1 July 2003.
- [9] Van Rijsbergen, C.J., 1979, “Information Retrieval Butterworths”, Department of Computing Science, University of Glasgow, second edition.
- [10] Yates, R. A. and B. Ribeiro-Neto, 1999, “Modern Information Retrieval”, Addison-Wesley.