



RAMA
UNIVERSITY

www.ramauniversity.ac.in

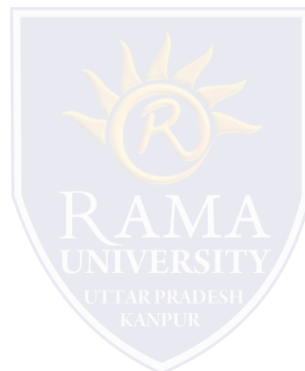
FACULTY OF ENGINEERING AND TECHNOLOGY

Soft Computing LECTURE -29

Umesh Kumar Gera
Assistant Professor
Computer Science & Engineering

OUTLINE

- **Application of soft computing**
- **Genetic algorithm based Internet Search Techniques**
- **References**
- **MCQ**



GENETIC ALGORITHM BASED INTERNET SEARCH TECHNIQUES

Problem Definition

Client-based searching agents can be agents for local search (which means that they search for documents linked to the input documents) and agents for non-local search (which can find documents that are not linked to the input documents).

Broadly defined, an agent is a program that can operate autonomously to accomplish unique tasks without direct human supervision.

The agent presented in this paper named *Tropical* is an agent for global search, written in the Java programming language.

For set of input documents, the agent finds and displays similar documents and the information regarding how similar they are. It uses services of following packages:

Spider [3] for document fetching

Agent [5] for Best First Search algorithm

Generator [6] for designing and communication with the database of topic sorted URLs.



Genetic Algorithm for Internet Search

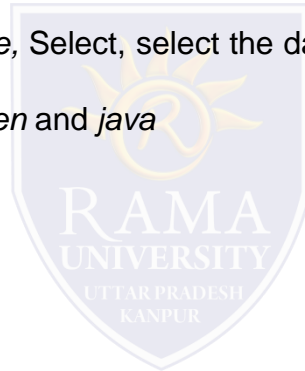
Algorithm performs following steps:

- I. User submits input set of documents like string of URLs.
- II. From the Web are fetched the input documents and parsed in order to extract keywords.
- III. From the Web are fetched documents that are pointed to by the hyperlinks in input documents and they form the current configuration set H.
- IV. From the database is randomly picked desired number of URLs that cover the same topic as do the input documents (this topic is specified by the user at the beginning of the program) and corresponding documents are fetched from the Web and inserted in the set H.
- V. Determining the documents from the set H that are most similar to the input ones by calculating Jaccard Score for each of them.
- VI. Inserting the most similar document in the output set and adding the documents linked to it in the set H.
- VII. Repeating steps 4-6 until desired number of output documents is obtained.

Some Relevant Details

Linking the database - just once before you first start the program

Start button in Windows environment, Settings, Control Panel, ODBC, User DSN Window, Add, Microsoft Access Driver, Finish, In the field Data Source Name put *URLbase*, Select, select the database *url.mdb* in the tropical directory, Advanced, in the fields for username and password type *mladen* and *java*



Starting the program

❑ In JDK 1.1.4 Package:

❑ `java tropical -dProxySet=true`

❑ `-dProxyHost=proksi`

❑ `-dProxyPort=port`

❑ *proksi* is IP address of proxy server you are using *port* number of this proxy server.

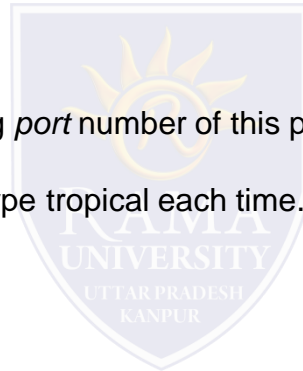
❑ you put this line in tropical.bat file and then just type tropical each time.

❑ Example of the command line is:

❑ `java tropical -dProxySet=true`

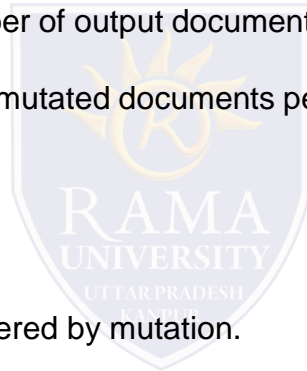
❑ `-dProxyHost=147.91.8.62`

❑ `-dProxyPort=8080`



Input parameters

- Input Set - field for input set of URLs. It is necessary that you enter either a set of URLs or a name of the file containing URLs or both.
- Input file - field for the name of the file with URLs (each in the new line)
- Number of outputs - field for the required number of output documents. Default number is 10.
- Database mutation - list for desired number of mutated documents per iteration. Default number is 0.
- Spatial mutation - not realized yet.
- Temporal mutation - not realized yet.
- Topic - list for choosing one or more topics covered by mutation.
- Comments - comments on/off field.



GENETIC ALGORITHM BASED INTERNET SEARCH TECHNIQUES

Output window

Program Flow - some program messages that keep you informed about program execution.

Result Set - result set of URLs.

Medium Jaccard's Score - medium Jaccard's Score for the result set.

Result of the program work can also be found in the file izlaz.html in the Tropical directory and can be viewed during program execution.



MULTIPLE CHOICE QUESTION

6. Once cosmids are inside the E.coli cells, they don't generate phage but are propagated as plasmids.

- a) True
- b) False

7. Which of the following is the correct method to check whether the DNA has entered into the cell or not in the case of cosmid?

- a) If a transformation has taken place turbid plaques are formed
- b) If a transformation has taken place clear plaques are formed
- c) If a transformation has taken place, it can be confirmed via ampicillin resistance
- d) If a transformation has taken place, it can be confirmed if forms plaques and is ampicillin resistant also

8. Which size of the insert is accepted by the cosmids?

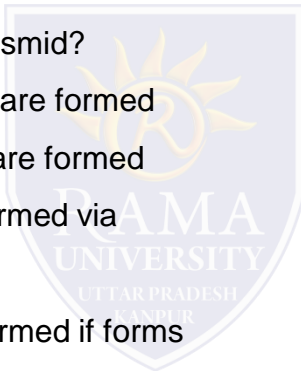
- a) 10-20 kbp
- b) 35-45 kbp
- c) 50-60 kbp
- d) 100-120 kbp

9. What happens once the cosmid enters the E. coli cells?

- a) There is strict size selection inside E. coli cells
- b) Partial deletion may take place
- c) The tendency of deletion may be increased by using low copy number
- d) The tendency of deletion can't be altered

10. If colEI derived origin of replication is replaced by the origin of replication of F plasmid, it is called as:

- a) phagemid
- b) F cosmid
- c) plasmid
- d) fosmid



REFERENCES

- ❑ https://shodhganga.inflibnet.ac.in/bitstream/10603/41504/12/12_chapter%202.pdf
- ❑ <https://rti.etf.bg.ac.rs/rti/ebi/dipl/jelena/paper.htm>

