

Disco Web



*Using link analysis to discover web communities
and improve search engine retrieval*

**Apostolos Gerasoulis, Brian D. Davison,
Konstantinos Kleisouris, Yingfang Lu, Hyun-ju Seo,
Wei Wang, Baohua Wu**

**Department of Computer Science
Rutgers University**

Mathematics

A: Connectivity Matrix

a: authorities; h: hubs

$$\begin{bmatrix} a \\ h \end{bmatrix}^{k+1} = \begin{bmatrix} 0 & A \\ A^t & 0 \end{bmatrix} \begin{bmatrix} a \\ h \end{bmatrix}^k$$

k=1 => "edge counting"



DiscoWeb Approach



- **Select initial web pages via text analysis (the results of a search engine query)**
- **Use link analysis to re-rank pages based on influence in graph**
- **Calculate multiple eigenvectors, each representing a community ranking**
- **More general method than CLEVER (IBM/Cornell) and Google (Stanford)**



DiscoWeb Results



- **DiscoWeb can retrieve top-rated sites (as listed by Infoseek's Directory)**
- **DiscoWeb also suggests other good sites (not included in Infoseek)**
- **DiscoWeb ranking is performed automatically, without human guidance**

Current Work



Goal: Fast, scalable search engine to provide large datasets and local content for link analysis

- Enable fast analysis
- Enable analysis of large graphs

Next tasks: distributed design, parallel solver, robust public version

DiscoWeb Interface

DiscoWeb!

New search: Display:

[art](#) [automotive](#) [business](#) [careers](#) [entertainment](#) [family](#)
[health](#) [kids](#) [money](#) [people](#) [science](#) [shopping](#)
[spirit](#) [sports](#) [test](#) [travel](#)

search results from DiscoWeb

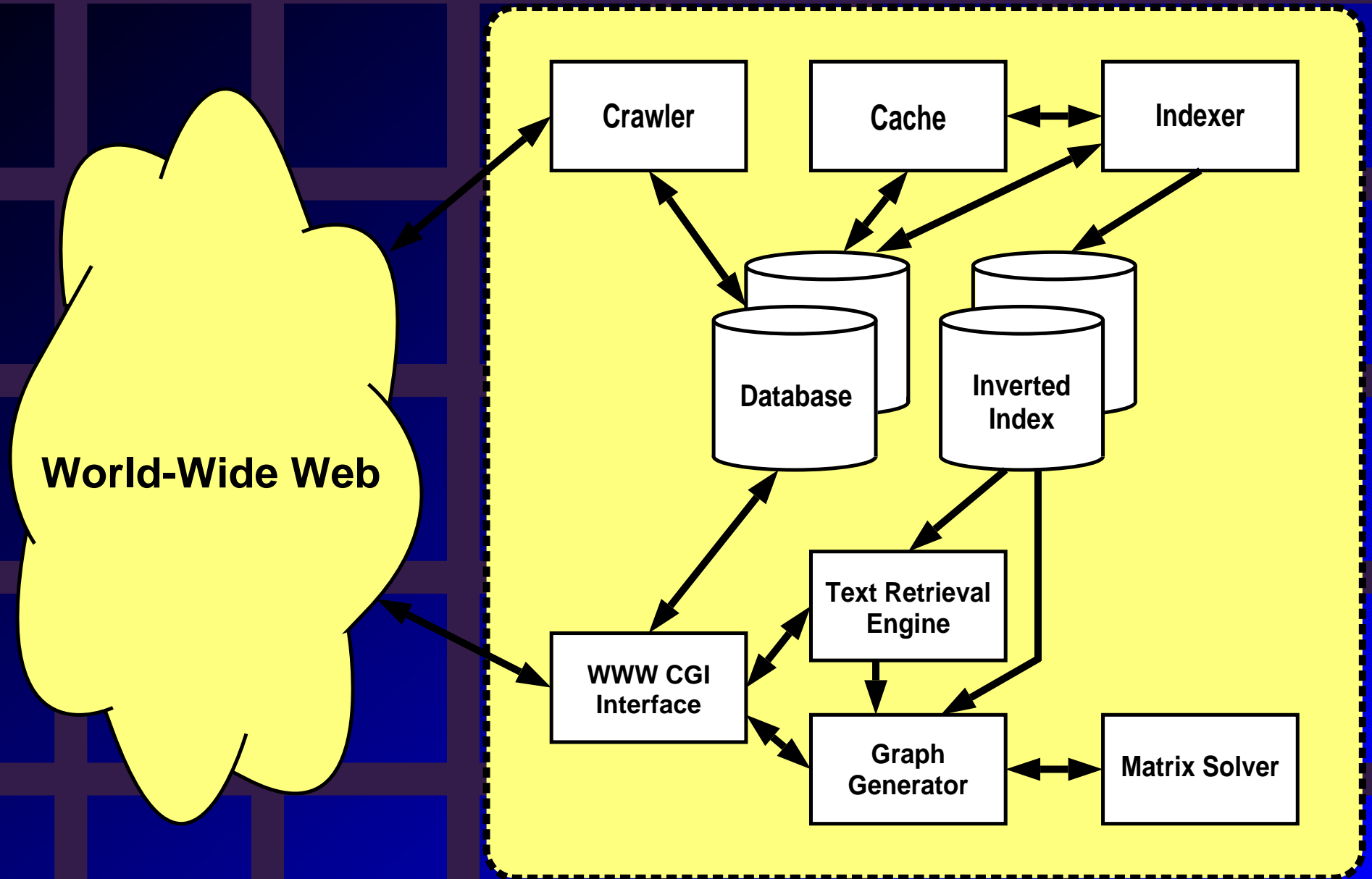
(Total : 7390)

Keywords : **engine**


Matches 1 - 10

- 1. Title :** [Alaska Search Engine : Canada Search Engine \[ALCANSEEK.COM\]](#)
Score: 5.0000 **URL :** <http://www.alcanseek.com/employment/publicate.htm>
Description: Alaska & Canada's official search engine. Search all of Alaska and Canada for travel, sports, business, news, entertainment and more....
- 2. Title :** [Alaska Search Engine : Canada Search Engine \[ALCANSEEK.COM\]](#)
Score: 5.0000 **URL :** <http://www.alcanseek.com/computersinternet/saleservice.htm>
Description: Alaska & Canada's official search engine. Search all of Alaska and Canada for travel, sports, business, news, entertainment and more....
- 3. Title :** [* BikeCrawler * Search Engine for the Cycling World * Bicycle Search Engine](#)
Score: 5.0000 **URL :** <http://www.bikecrawler.com/>
Description: A search engine targeted toward the bicycling industry. The BikeCrawler Search Engine contains search categories for manufacturers, tour companies, bicycle clubs

DiscoWeb Infrastructure



Clusters

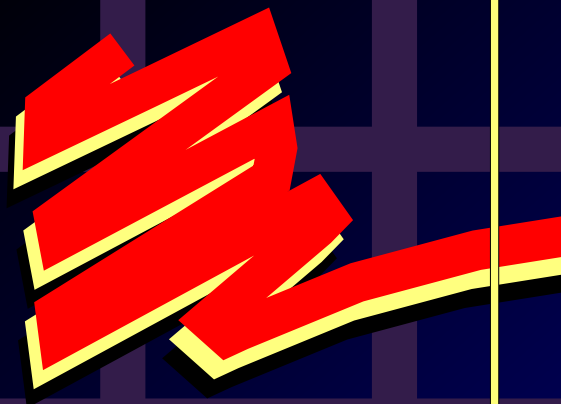
- 
- Link analysis produces clusters corresponding to web communities.
 - Each eigenvalue gives the weight of the cluster.
 - Values in each eigenvector rank pages in that community.



Example of Family



- **There are many different communities represented.**
- **Communities include genealogy, family health, parenthood, Christian family values, family care givers.**



"Family"

***Search
Results***

DiscoWeb

Netscape: Eigenvalue51

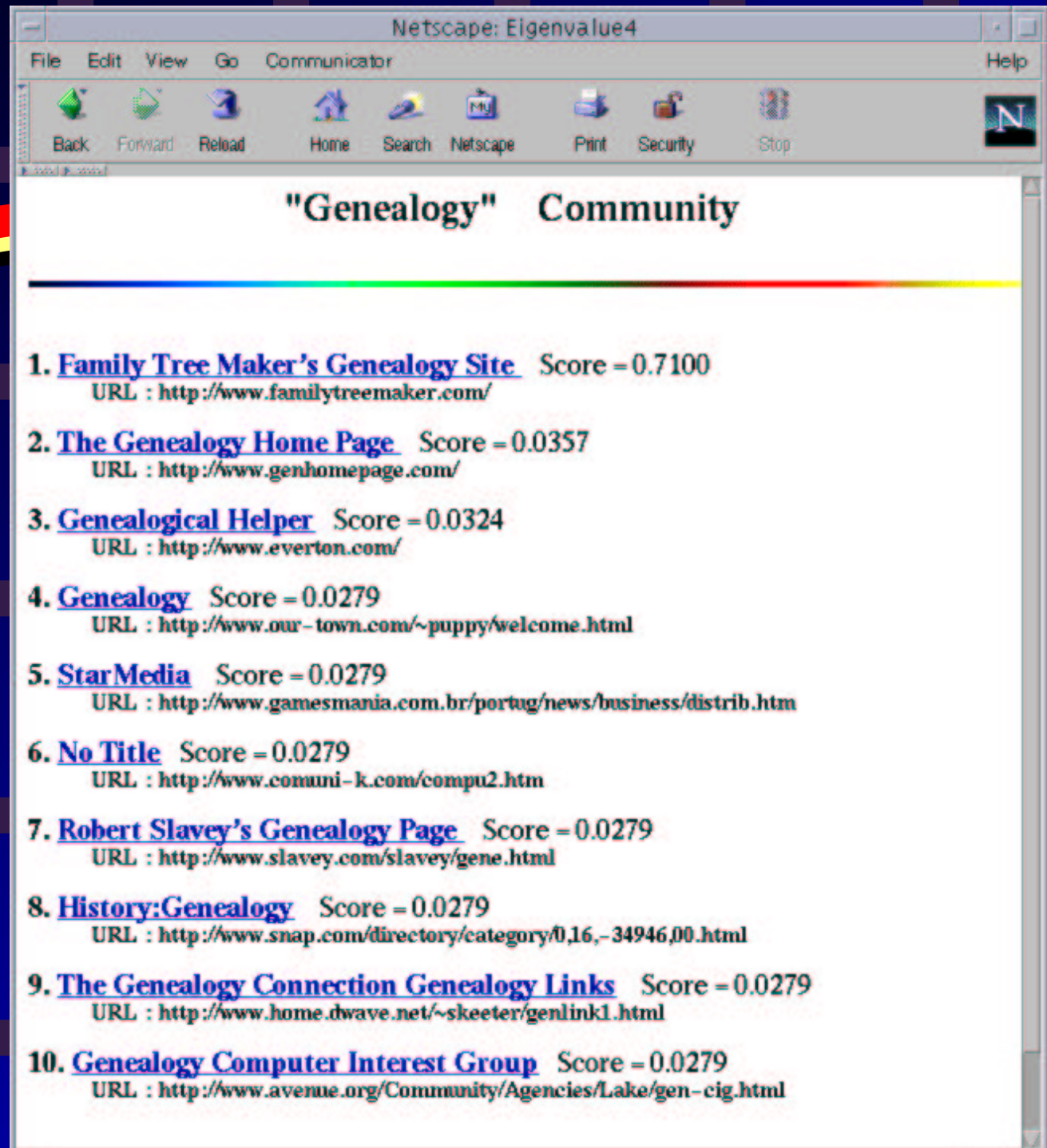
File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Search Results for "family"

1. [Genealogical Helper](http://www.everton.com/) Score = 0.7979
URL : <http://www.everton.com/>
2. [The Genealogy Home Page](http://www.genhomepage.com/) Score = 0.7845
URL : <http://www.genhomepage.com/>
3. [Family Tree Maker's Genealogy Site](http://www.familytreemaker.com/) Score = 0.7518
URL : <http://www.familytreemaker.com/>
4. [Focus on the Family with Dr. James Dobson](http://www.family.org/) Score = 0.7115
URL : <http://www.family.org/>
5. [American Academy of Family Physicians Home Page](http://www.aafp.org/) Score = 0.6613
URL : <http://www.aafp.org/>
6. [Welcome to HomeArts!](http://homearts.com/) Score = 0.4572
URL : <http://homearts.com/>
7. [Welcome to Parenting Q&A](http://www.parenting-qa.com/) Score = 0.4431
URL : <http://www.parenting-qa.com/>
8. [Family Caregiver Alliance](http://www.caregiver.org/) Score = 0.4371
URL : <http://www.caregiver.org/>
9. [Dove Family Approved Movies and Videos](http://www.dove.org/) Score = 0.4369
URL : <http://www.dove.org/>
10. [Crayola](http://www.crayola.com/) Score = 0.4293
URL : <http://www.crayola.com/>

"Genealogy"



The screenshot shows a Netscape browser window titled "Netscape: Eigenvalue4". The browser's menu bar includes "File", "Edit", "View", "Go", "Communicator", and "Help". The toolbar contains icons for "Back", "Forward", "Reload", "Home", "Search", "Netscape", "Print", "Security", and "Stop". The main content area displays the search results for the query "Genealogy".

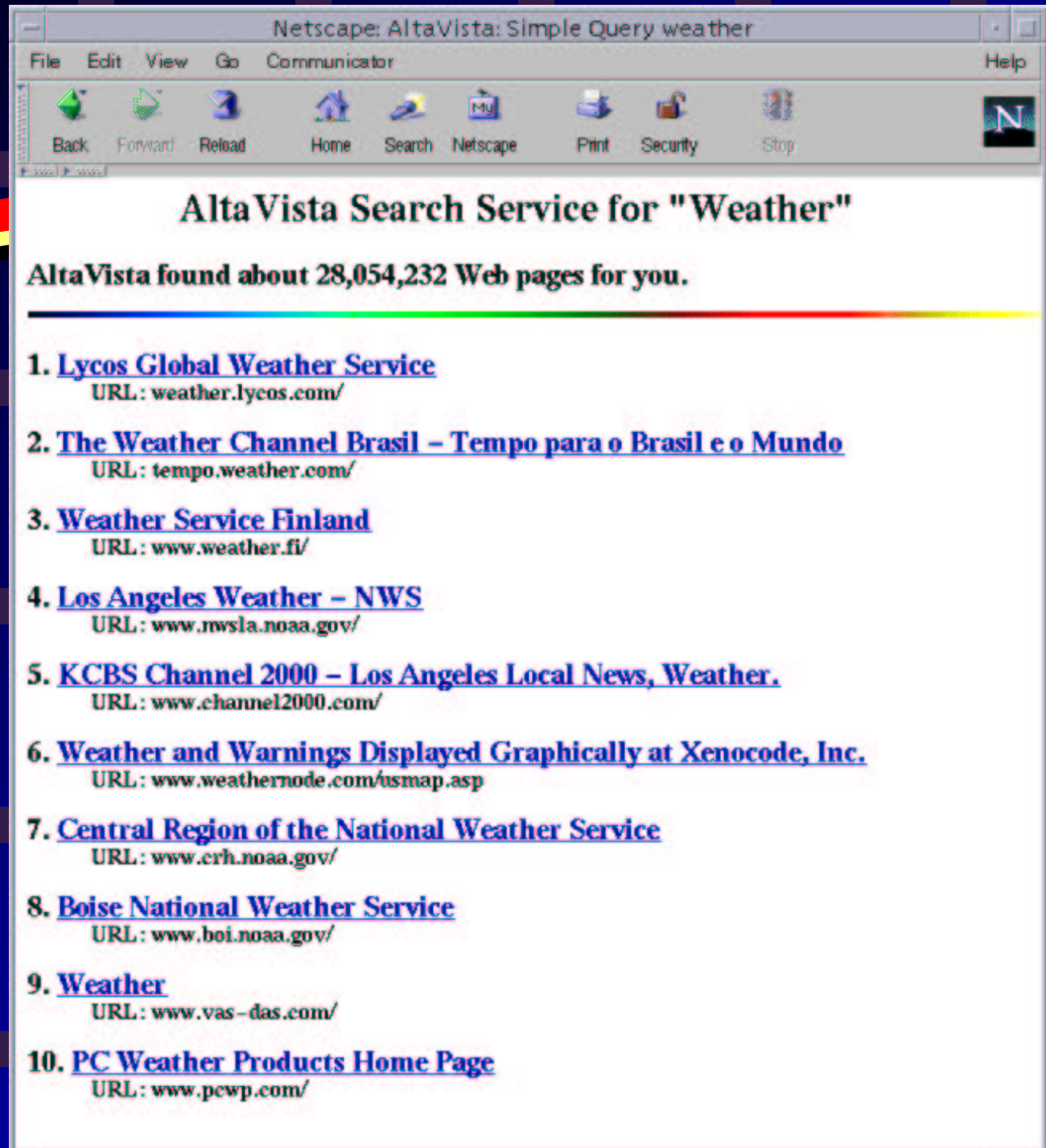
"Genealogy" Community

1. [Family Tree Maker's Genealogy Site](http://www.familytreemaker.com/) Score = 0.7100
URL : <http://www.familytreemaker.com/>
2. [The Genealogy Home Page](http://www.genhomepage.com/) Score = 0.0357
URL : <http://www.genhomepage.com/>
3. [Genealogical Helper](http://www.everton.com/) Score = 0.0324
URL : <http://www.everton.com/>
4. [Genealogy](http://www.our-town.com/~puppy/welcome.html) Score = 0.0279
URL : <http://www.our-town.com/~puppy/welcome.html>
5. [StarMedia](http://www.gamesmania.com.br/portug/news/business/distrib.htm) Score = 0.0279
URL : <http://www.gamesmania.com.br/portug/news/business/distrib.htm>
6. [No Title](http://www.comuni-k.com/compu2.htm) Score = 0.0279
URL : <http://www.comuni-k.com/compu2.htm>
7. [Robert Slavey's Genealogy Page](http://www.slavey.com/slavey/gene.html) Score = 0.0279
URL : <http://www.slavey.com/slavey/gene.html>
8. [History:Genealogy](http://www.snap.com/directory/category/0,16,-34946,00.html) Score = 0.0279
URL : <http://www.snap.com/directory/category/0,16,-34946,00.html>
9. [The Genealogy Connection Genealogy Links](http://www.home.dwave.net/~skeeter/genlink1.html) Score = 0.0279
URL : <http://www.home.dwave.net/~skeeter/genlink1.html>
10. [Genealogy Computer Interest Group](http://www.avenue.org/Community/Agencies/Lake/gen-cig.html) Score = 0.0279
URL : <http://www.avenue.org/Community/Agencies/Lake/gen-cig.html>

"Weather"

**Search
Engine
Results**

AltaVista



Netscape: AltaVista: Simple Query weather

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

AltaVista Search Service for "Weather"

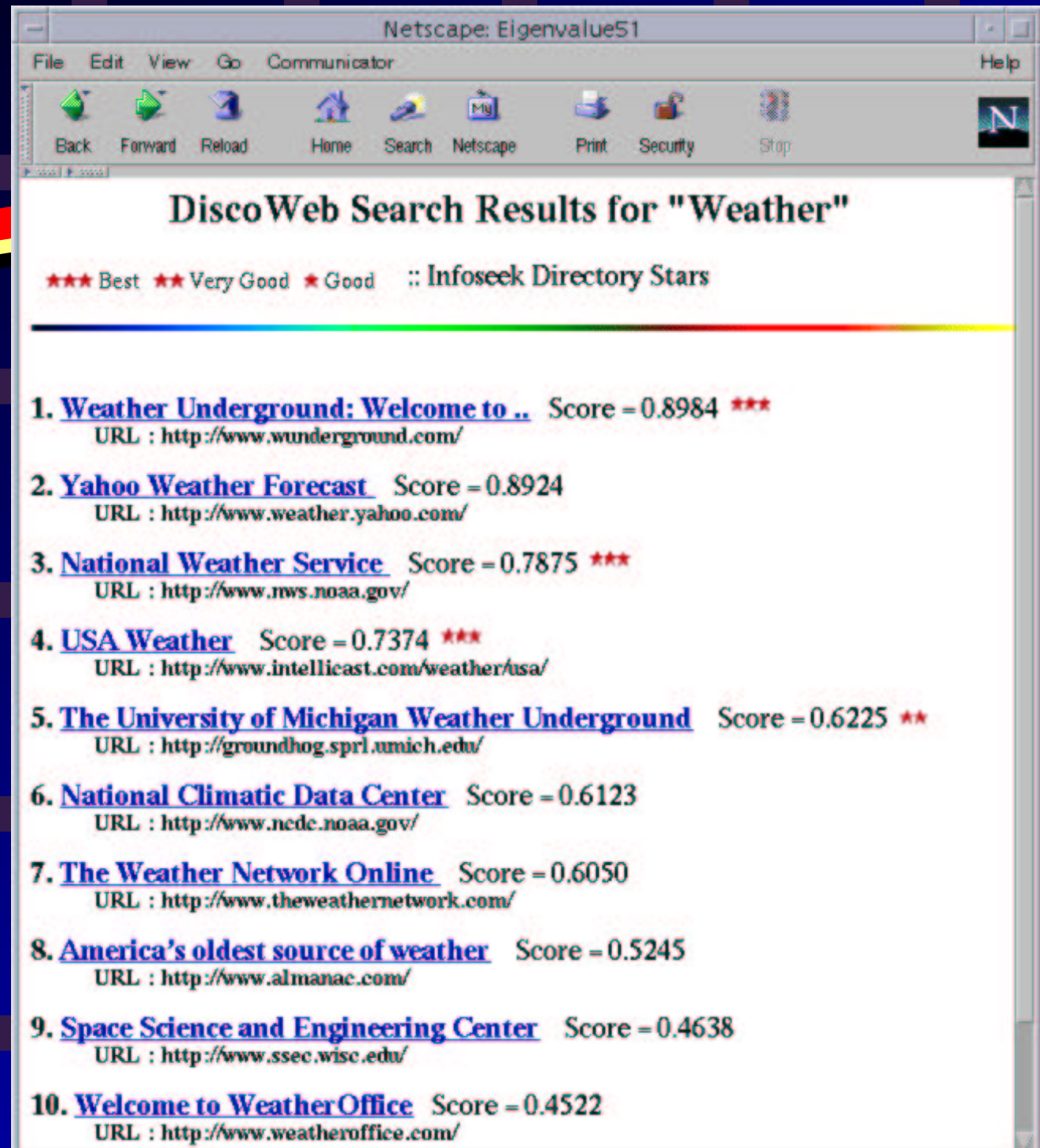
AltaVista found about 28,054,232 Web pages for you.

1. [Lycos Global Weather Service](http://weather.lycos.com/)
URL: weather.lycos.com/
2. [The Weather Channel Brasil - Tempo para o Brasil e o Mundo](http://tempo.weather.com/)
URL: tempo.weather.com/
3. [Weather Service Finland](http://www.weather.fi/)
URL: www.weather.fi/
4. [Los Angeles Weather - NWS](http://www.nwsla.noaa.gov/)
URL: www.nwsla.noaa.gov/
5. [KCBS Channel 2000 - Los Angeles Local News, Weather.](http://www.channel2000.com/)
URL: www.channel2000.com/
6. [Weather and Warnings Displayed Graphically at Xenocode, Inc.](http://www.weathernode.com/usmap.asp)
URL: www.weathernode.com/usmap.asp
7. [Central Region of the National Weather Service](http://www.erh.noaa.gov/)
URL: www.erh.noaa.gov/
8. [Boise National Weather Service](http://www.boi.noaa.gov/)
URL: www.boi.noaa.gov/
9. [Weather](http://www.vas-das.com/)
URL: www.vas-das.com/
10. [PC Weather Products Home Page](http://www.pcwp.com/)
URL: www.pcwp.com/

"Weather"

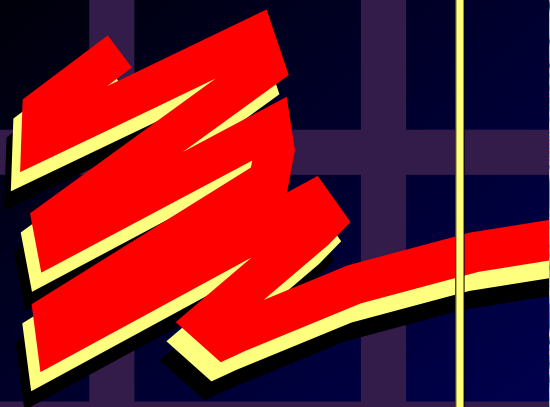
**Link
Analysis
Results**

DiscoWeb



The screenshot shows a Netscape browser window titled "Netscape: Eigenvalue51". The address bar is empty. The menu bar includes "File", "Edit", "View", "Go", "Communicator", and "Help". The toolbar contains icons for "Back", "Forward", "Reload", "Home", "Search", "Netscape", "Print", "Security", and "Stop". The main content area displays the title "DiscoWeb Search Results for 'Weather'" and a legend for "Infoseek Directory Stars" with three star ratings: "★★★ Best", "★★ Very Good", and "★ Good". A list of 10 search results follows, each with a score and a star rating. A red arrow points from the "Weather" text on the left towards the search results.

Rank	Search Result	Score	Rating
1.	Weather Underground: Welcome to ..	0.8984	★★★
2.	Yahoo Weather Forecast	0.8924	
3.	National Weather Service	0.7875	★★★
4.	USA Weather	0.7374	★★★
5.	The University of Michigan Weather Underground	0.6225	★★
6.	National Climatic Data Center	0.6123	
7.	The Weather Network Online	0.6050	
8.	America's oldest source of weather	0.5245	
9.	Space Science and Engineering Center	0.4638	
10.	Welcome to WeatherOffice	0.4522	



"Java"

**Search
Engine
Results**

AltaVista

Netscape: AltaVista: Simple Query Java

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

AltaVista Search Service for "Java"

AltaVista found about 18,254,865 Web pages for you.

1. [hav. Software – Neural Nets to Internet – Java et.al.](http://www.hav.com/)
URL: www.hav.com/
2. [WingSoft – One Stop Shop for Java Tools](http://www.wingsoft.com/)
URL: www.wingsoft.com/
3. [Caf151soft: A Java Software Development Company](http://www.cafesoft.com/)
URL: www.cafesoft.com/
4. [CORBA, COM, RPC, and Java development](http://www.nouveau-corba.com/)
URL: www.nouveau-corba.com/
5. [Natural Intelligence – Java, Oracle..](http://www.natural.com/)
URL: www.natural.com/
6. [Servlet Central: The Server-side Java Magazine](http://www.servletcentral.com/)
URL: www.servletcentral.com/
7. [Munica CGI SQL JAVA and Database Application Center](http://www.servletcentral.com/)
URL: www.servletcentral.com/
8. [JC VRML & JAVA Consulting: 3D Sports for the Web](http://www.technowin.com/)
URL: www.technowin.com/
9. [WebStrike! – an online java game](http://www.webstrike.com/)
URL: www.webstrike.com/
10. [Java and Web Services](http://www.jaws.umn.edu/)
URL: www.jaws.umn.edu/



"Java"

**Link
Analysis
Results**

DiscoWeb

Netscape: Infoseek: Java

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

DiscoWeb Search Results for "Java"

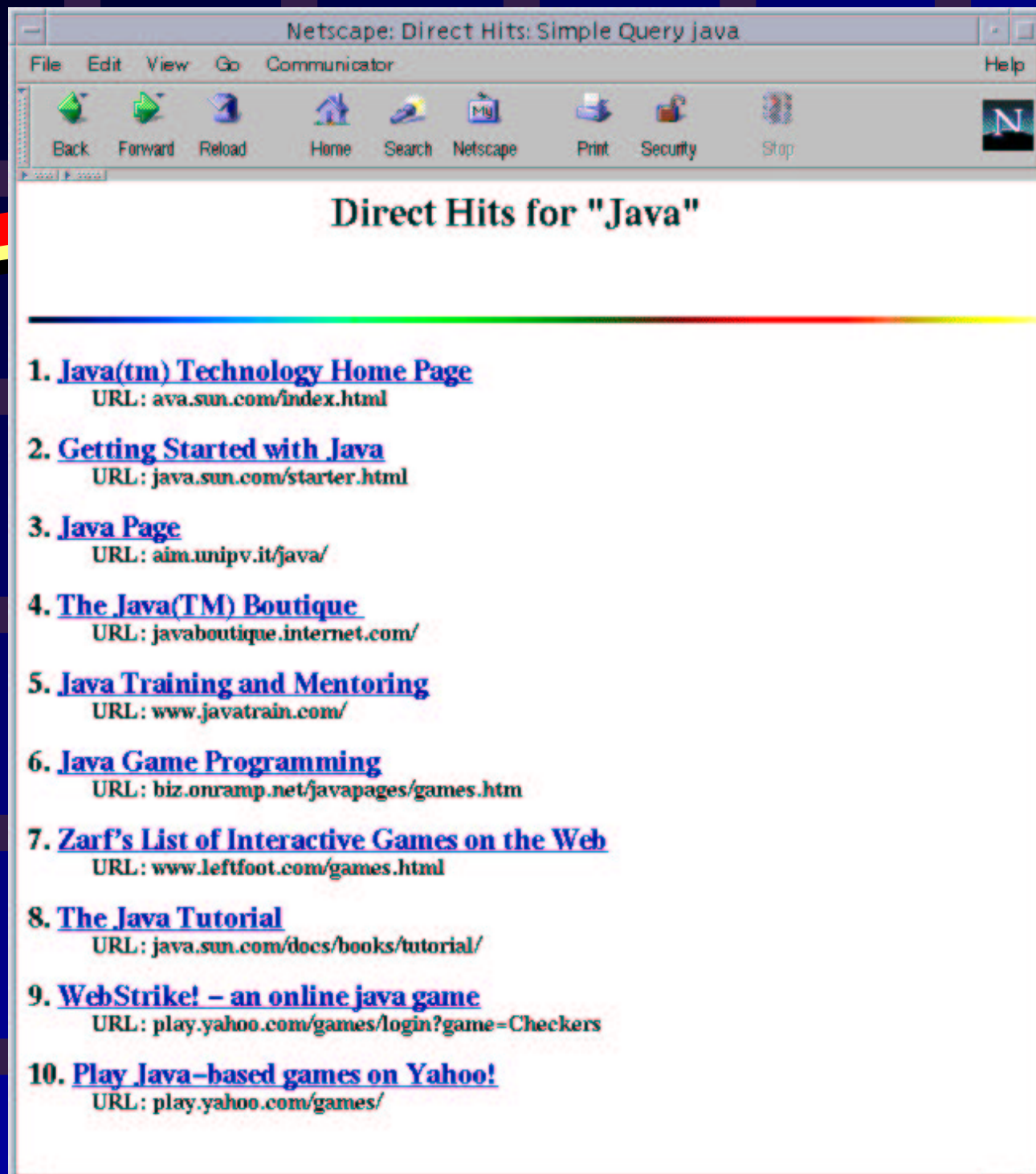
*** Best ** Very Good * Good :: Infoseek Directory Stars

- 1. [Java Applet Rating Service \(IARS\)](http://www.jars.com) Score = 1.0837 ***
URL: www.jars.com**
- 2. [Java Linux](http://www.blackdown.org/java-linux.html) Score = 1.0027
URL: www.blackdown.org/java-linux.html**
- 3. [Gamelan: Earthweb's Java Directory](http://www.gamelan.com/vx/) Score = 0.9196 ***
URL: www.gamelan.com/vx/**
- 4. [Java\(TM\) Technology Home Page](http://java.sun.com/) Score = 0.8526 **
URL: java.sun.com/**
- 5. [CodeWarrior](http://www.metrowerks.com/) Score = 0.7082
URL: www.metrowerks.com/**
- 6. [The Java Apache Project](http://java.apache.org/) Score = 0.5893
URL: java.apache.org/**
- 7. [Java Repository: for Java Programmers](http://java.wivi.uni-frankfurt.de/) Score = 0.5891 ***
URL: java.wivi.uni-frankfurt.de/**
- 8. [Front Page -- www.arachnoid.com](http://www.arachnoid.com) Score = 0.5665
URL: [www.arachnoid.com/](http://www.arachnoid.com)**
- 9. [Artistic Reality](http://www.artreality.com) Score = 0.5329
URL: www.artreality.com**
- 10. [Java Lobby Homepage](http://www.javalobby.org/) Score = 0.5310 **
URL: www.javalobby.org/**

"Java"

**Direct Hits
from**

HotBot



The screenshot shows a Netscape browser window titled "Netscape: Direct Hits: Simple Query java". The browser interface includes a menu bar (File, Edit, View, Go, Communicator, Help) and a toolbar with icons for Back, Forward, Reload, Home, Search, Netscape, Print, Security, and Stop. The main content area displays the search results for "Java".

Direct Hits for "Java"

1. [Java\(tm\) Technology Home Page](http://ava.sun.com/index.html)
URL: ava.sun.com/index.html
2. [Getting Started with Java](http://java.sun.com/starter.html)
URL: java.sun.com/starter.html
3. [Java Page](http://aim.unipv.it/java/)
URL: aim.unipv.it/java/
4. [The Java\(TM\) Boutique](http://javaboutique.internet.com/)
URL: javaboutique.internet.com/
5. [Java Training and Mentoring](http://www.javatrain.com/)
URL: www.javatrain.com/
6. [Java Game Programming](http://biz.onramp.net/javapages/games.htm)
URL: biz.onramp.net/javapages/games.htm
7. [Zarf's List of Interactive Games on the Web](http://www.leftfoot.com/games.html)
URL: www.leftfoot.com/games.html
8. [The Java Tutorial](http://java.sun.com/docs/books/tutorial/)
URL: java.sun.com/docs/books/tutorial/
9. [WebStrike! – an online java game](http://play.yahoo.com/games/login?game=Checkers)
URL: play.yahoo.com/games/login?game=Checkers
10. [Play Java-based games on Yahoo!](http://play.yahoo.com/games/)
URL: play.yahoo.com/games/



"Java"
**Search
Engine
Results**
HotBot

Netscape: Hotbot: Simple Query java

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Hotbot Search Service for "Java"

Hotbot found about 1,926,360 Web pages for you.

1. [Java Technology Home Page](http://www.java.sun.com/)
URL: www.java.sun.com/
2. [Gamelan](http://www.gamelan.com/)
URL: www.gamelan.com/
3. [MediaBuilder - Java Animation](http://deutsch.mediabuilder.com/softwaregraphicsjava.html)
URL: deutsch.mediabuilder.com/softwaregraphicsjava.html
4. [ComSpace.Com](http://www.comspace.com/)
URL: www.comspace.com/
5. [Digital Cats' Java\(tm\)](http://www.javacats.com/)
URL: www.javacats.com/
6. [Sun Microsystems' Java Platform](http://ac030.osu.cz/java/jdk/docs/api/packages.html)
URL: ac030.osu.cz/java/jdk/docs/api/packages.html
7. [Sun Microsystems' Java Platform](http://www2.cs.utah.edu/classes/cs3510/jdk1.1.7B/docs/api/packages.html)
URL: www2.cs.utah.edu/classes/cs3510/jdk1.1.7B/docs/api/packages.html
8. [Netpedia-->The Solution for Web Developers](http://netpedia.com/)
URL: netpedia.com/
9. [programmers' and developers' resources](http://www-a.developer.com/)
URL: www-a.developer.com/
10. [Package Summary](http://service.felk.cvut.cz/doc/java/share/jdk1.2beta3/docs/api/packages.html)
URL: service.felk.cvut.cz/doc/java/share/jdk1.2beta3/docs/api/packages.html