

# Information Search 3: UNITERNET

Farzaneh Aminpour, PhD.
aminpour@tums.ac.ir
Tehran University of Medical Sciences

# الانتان الانتان الدوج على المادية على المادية **Workshop Objectives**

- Internet Search Methods
- Web Directories
- Web Search Engines
- Googling: Basic, Advanced, Top Tips
- Scholar Google: Search, Citation
- Medical Search Engines
- Medical Metasites
- Accreditation of Medial Websites

- - Choose proper **keywords**
  - Use related **search tools**
  - التنشارات Use advanced search techniques
  - Design the best search strategy
  - Enter your search query
  - Retrieve proper Abstracts
  - Retrieve proper Fulltexts
  - **Search in Information Databases**
  - 9. Search in Electronic Journal Collections
  - o. Search in other web resources, Googling



#### Internet

- در ده الانتار الن Internet is a global system of interconnected computer networks.
- The terms Internet and World Wide Web are often used in everyday speech without much distinction. However, they are not one and the same.



- Web is one of the services that runs on the Internet.
- Web can be thought of as an application running on the Internet.
- Web is a collection of text documents and other resources, linked by hyperlinks and URLs, usually accessed by web browsers from web servers.



- The most useful way of thinking of the Internet is as a huge library, with each web page representing a book in that library.
- More than 200 million Websites
- Approximately 3 milliard web pages
- Doubling time of web documents is about 8 months



- peer into the holdings of a library or archive
- search through databases hosted there
- connect with other researchers with similar interests
- share our discoveries with others
- link to other related sites
- explore assorted resources on favorite topics



- In 1945, Vannevar Bush described **Memex** (a portmanteau of **memory** and **index**) in his article "As We May Think" in "The Atlantic Monthly" as a device in which individuals would compress and store all of their books, records, and communications, "mechanized so that it may be consulted with exceeding speed and flexibility."
- Memex influenced the development of hypertext systems and the creation of the World Wide Web and personal knowledge base software.

м

■ In the May 1970 issue of "Popular Science"
magazine Arthur C Clair satellites would some day "bring the accumulated knowledge of the world to your fingertips" using a console that would combine the functionality of the photocopier, telephone, television and a small computer, allowing data transfer and video conferencing around the globe.

- .
  - In March 1989, Tim Berners-Lee wrote a proposal that referenced ENQUIRE, a database and software project he had built in 1980, and described a more elaborate information management system.
  - On 12 November 1990, Tim Berners-Lee With help from Robert Cailliau, published a more formal proposal to build a "Hypertext project" called "WorldWideWeb" as a "web" of "hypertext documents" to be viewed by "browsers" using a client—server architecture.

м

 On 6 August 1991, Tim Berners-Lee posted a short summary of " Web project on the alt.hypertext newsgroup. This date also marked the debut of the Web as a publicly available service on the Internet.

## Two ways to find information و علم سنبه on the web

- 1. Web Directories
- 2. Web Search Engines المراجعة الم



#### **Directories**

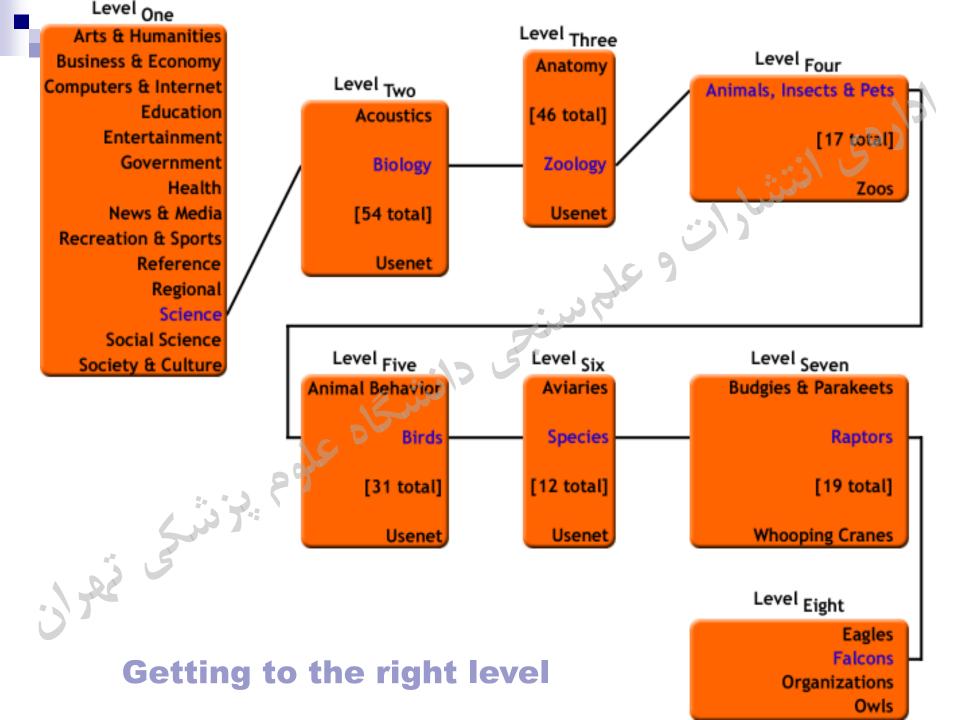
■ Web directories require *people* to view the individual Web site and determine its placement into a subject classification scheme or taxonomy. Once done, certain keywords associated with those sites can be used for searching the directory's data banks to find Web sites of interest.



- For searches that are easily classified, the search directories tend to provide the most consistent and well-clustered results. This advantage is generally limited solely to those classification areas already used in the taxonomy by that service.
- Yahoo, for example, has about 2,000 classifications in its current taxonomy. When a given classification level reaches 1,000 site listings or so, the Yahoo staff split the category into one or more subcategories.



- If a given topic area has *not been* specifically classified by the search directories, finding any related information on that topic is made more difficult.
- lack of coverage because of the cost and time in individually assigning sites to categories.



### **Examples of Web Directories**

- Yahoo
- Open Directory
- LookSmart
- Librarian Index
- Infomine
- Academic Info
- About.com

http://dir.yahoo.com

http://dmoz.org

http://www.looksmart.com

http://lii.org

http://infomine.ucr.edu

http://www.academicinfo.net

http://www.about.com







#### Search Engines

There are more than 2500 search services presently on the web.





- Search engines use Spiders (Crawlers) or Robots to go out and retrieve individual web pages or documents.
- Then they will make index files.



- A search engine operates, in the following order
  - □ Web crawling
  - □ Indexing
  - □ Searching



- 1. Because many searches are not very well defined, indexes will often be the *best* starting point.
- 2. Indexes, as they cover most (or at least more) words on a given page will offer a richer list of returns.
- 3. Indexes are usually larger because of the much lower overhead in adding pages to the search engine (more sensitivity)



 The flexibility of indexing every word to give users complete search control is now creating a different kind of problem: too many results (lack of specificity).

Since the user is limited to viewing potential sites one-by-one, clearly too many results can be a greater problem than too few.

#### Contents of a web page

- 1. Title: It is viewable in the blue bar if the webpage.
- 2. Description: a type of metatag which provides a short, summary description provided by the document designer; not viewable on the actual page; this is frequently the description of the document shown on the documents listings by the search engines that use metatags.
- 3. Keywords: another type of metatag consisting of a listing of keywords that the document designer wants search engines to use to identify the document. These too, are **not viewable** on the actual page
- Body: the actual viewable content of the document.



A Web page, or document, can contain various kinds of content, some of which is not shown when you view the document in your browser: like its Source.



- Search engines may index all or some of these content fields when stories document on their databases.
- Over time, engines have tended to index fewer words and fields.
- Then, using algorithms that differ substantially from engine to engine, when a search query is evaluated by that engine its listing of document results is presented in order of relevance.



#### The Differences

■ Because of the differences in degree of indexing and algorithms used, the same document listed on different search engines can appear at a much higher or lower ranking (order of presentation) than on other engines.



Highly variable from engine to engine, four factors tend to influence greatly the ranking of a document in a given query:

#### Ranking of documents

- 1. Order a keyword term appears: keyword terms that appear sooner in the document's listing or index tend to be ranked higher
- Frequency of keyword term: keywords that appear multiple times in a document's index tend to be ranked higher
- 3. Occurrence of keyword in the title: keywords that appear in the document's title, or perhaps metatag description or keyword description fields, can be given higher weight than terms only in the document body
- 4. Rare, or less frequent, keywords: rare or unusual keywords that do not appear as frequently in the engine's index database are often ranked more highly than common terms or keywords.

# But none of them come close to indexing the entire Web...!

- Content of Adobe PDF and formatted files
- The content in sites requiring a log in
- Intranets; pages not linked from anywhere else
- Commercial resources with domain limitations
- Sites that use a robots.txt file to keep files and/or directories off limits
- Non-Web resources

### Examples of search engines

- AltaVista
- Excite
- FAST
- Google
- HotBot
- Northern Light

http://www.altavista.com

http://www.excite.com

http://www.alltheweb.com

http://www.google.com

http://www.hotbot.com

http://www.northernlight.com







### Health specific search Je 9 (3) engines

- Medstory
  - □ http://www.medstory.com/
- Eheath Sites
  - □ http://www.ehealthsites.com/
- Med Explorer
  - □ http://www.medexplorer.com/
- Mayo Clinic Health Oasis
  - http://www.mayohealth.org
- Medical World Search
  - http://www.mwsearch.com

## Health specific search المناسلة المناسلة المناسلة علم المناسلة engines

- TextMed
  - □ <a href="http://www.textmed.com/">http://www.textmed.com/</a>
- OnHealth
  - □ http://www.onhealth.com
- MedHunt
  - □ <a href="http://www.hon.ch/MedHunt/">http://www.hon.ch/MedHunt/</a>
- Md Tools
  - http://www.mdtool.com/
- **Nurse Web Search** 
  - □ <a href="http://www.nursewebsearch.com/">http://www.nursewebsearch.com/</a>



#### Multiple Search

Understanding how search engines work and combining multiple search engines in a given search can increase the likelihood of finding the information desired by a factor of 3.5 or more.

#### Search engines for search الانتنا وه التنس engines

- Search Engine Colossus
  - □ http://www.searchenginecolossus.com/
- Search Engine Watch
  - □ http://searchenginewatch.com
- Search Engine Showdown

  □ http://www.searchengineshowdown.com/



### Popular Meta Search Engines

- Dogpile
- Metacrawler
- Search.com
- Inference FIND
- Internet Sleuth
- Mamma

http://www.dogpile.com

http://www.metacrawler.com

http://www.search.com

http://www.infind.com

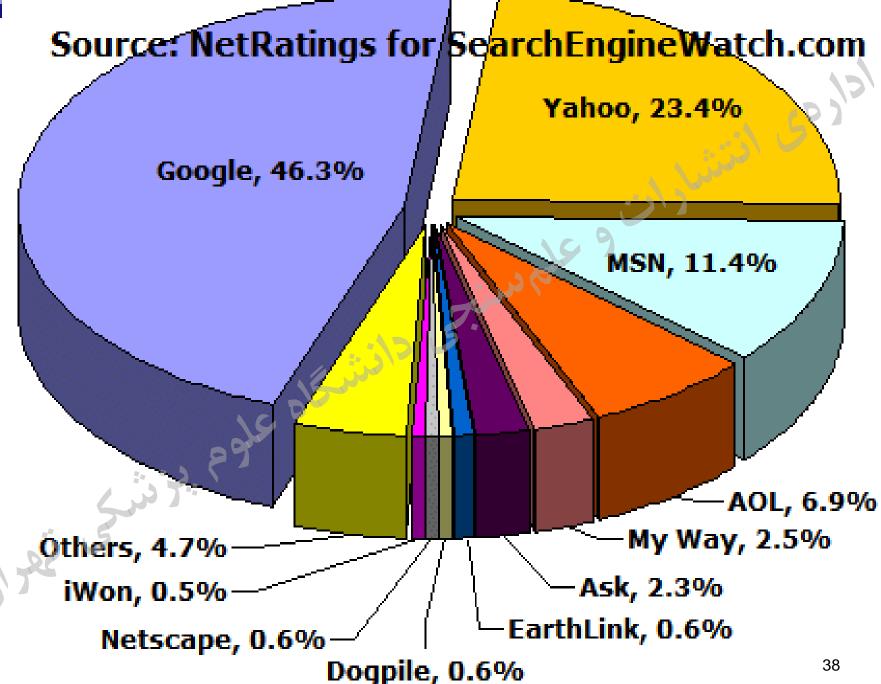
http://www.isleuth.com

http://www.mamma.com



- Search engines are machine-based and index words or terms in internet documents.
- Directories are human-based and classify web documents or locations into an arbitrary subject classification scheme or taxonomy.







In September of 1998, Google Inc. opened its door in Menlo Park, CA.

(http://www.google.com/intl/en/corporate/history.html#1995)





## **About Google**

- "Google" is a play on the word "googol" which is the mathematical term for 1 followed by 100 zeros.
- It reflects Google's mission is to organize the world's information and make it universally accessible and useful.

### Google: A Unique Company

4 Unique Company







**Sergey Brin** 

- ✓ The interface is clear and simple.
- ✓ Pages load instantly.
- ✓ No pop-up ads allowed.

## Why do we love Google?

- Size and scope: Now indexing over 20 billion web pages (conservative estimate).
- Relevance of Results: Page Rank
- **Diversity of Search:** Image, News, Book Search, Scholar, Blog Search, Finance, Froogle, Video . . . and much, much more.
- Other Tools: Google Maps, Gmail, Calendar, Docs & Spreadsheets . . .

Googling: Junii Googling: Juni



- An "I'm Feeling Lucky" search means you spend less time searching for web pages and more time looking at them.
- "I'm Feeling Lucky" button on the homepage takes you directly to the first webpage that returns for your query. When you use the feature, you won't see the other search results at all.



- Use quotes to search for an exact word or set of words.
  - [ "hospital management" ]
- Only use this if you're looking for a very precise word or phrase, because otherwise you could be excluding helpful results by mistake.

### Exclude a word -query

- النتنا ده ا Add a dash (-) before a word or site to exclude all results that include that word.
  - [ hepatitis –india ]
- You can also exclude results based on other operators, like excluding all results from a specific site.
  - [google -site:wikipedia.org]

# Include similar words ~query

Normally, synonyms might replace some words in your original query. Add a tilde sign (~) immediately in front of a word to search for that word as well as even more synonyms.

[ ~food facts ] includes results for "nutrition facts"



- عرابة المنابع المارية Find pages in the format you prefer.

# Search within a site/domain site:query

- Include "site:" to search for information within a single website. [vaccination site:who.int]
- Also search within a specific top-level domain like .org or .edu or country toplevel domain like .de or .jp. [vaccination site:.gov]

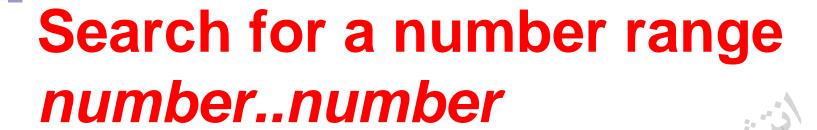
# Include a "fill in the blank" query \* query

Use an asterisk (\*) within a query as a placeholder for any unknown or "wildcard" terms. Use with quotation marks to find variations of that exact phrase or to remember words in the middle of a phrase.

[ "a \* saved is a \* earned" ]



- در دی انتشار If you want to search for pages that may have just one of several words, include OR (capitalized) between the words. Without the OR, your results would typically show only pages that match both terms. [ AIDS OR HIV]
- Enclose phrases in quotes to search for either one of several phrases.
  - ["cardiac arrest" OR "myocardial infarction"]



- Separate numbers by two periods (with no spaces) to see results that contain numbers in a given range of things like dates, prices, and measurements.
  [ ambu bag \$50..\$100]
- Use only one number with two periods for an upper maximum or a lower minimum.
   [ambu bag ..\$100]



- To find pages that link to a site or page like www.google.com, use [link:google.com].
- You can also search for links to specific pages like [link:google.com/images].



### **Advanced Search**

- For certain complex searches, you might want more control over the results that you see. Advanced Search makes the results more precise and more useful.
- Click on the gear icon in the top right corner of the search results page then clicking Advanced search.



### **Exceptions**

- انتنا ده ا Most search rules have exceptions to accommodate the ways that people commonly search.
- For example, Google will show calculator results for the query [ 34 \* 87 ] rather than interpreting the asterisk as the "fill in the blanks" operator.

## Find definitions

- finitio.

  fine:heamopi.

  define:asthma Find definition of a word or a phrase
  - □ define:heamophilia

# Google works like a calculator, too 5+2\*2 (addition, multiplication)

- 2<sup>3</sup> (exponent)(shift+6 keys)
- Sqrt100 (square root) المناعدة المعادد

### v

### Google as a Convertor

#### Google could convert scales & currency units

- 13.75 USD in Euro (convert \$13.75 in U.S. dollars to Euro)
- 220 Eur in JPY (convert 220 Euro in Japanese Yen)
- 27 miles in kilometers (convert 27 miles to kilometer)
- 2.3 pound in kilogram (convert 2.3 pound to kilogram)

# Google Translate http://translate.google.com/



Translate a Web Page Translate an entire web page

http://

English to French

Translate

Translate



- Example: Diabetes
- Why important: actions (verbs), modifiers (adjectives, and adverbs), and conjunctions are either "thrown away" by search engines or too variable to be useful



### Use 6 to 8 words in query

- Example: Diabetes, mellitus, neuropathy, sensory, treatment, biguanide
- Why important: more keywords chosen at appropriate level, can reduce the universe of possible documents returned by 99%



- Example: "diabetes mellitus" (treatment OR therapy)
- Why important: triangulating on multiple query concepts, narrows and targets results, generally by more than 100 to 1000



- Example: treatment OR therapy
- Why important: cover the likely different ways a concept can be described.
   Generally avoid OR in other cases



- Example: ("diabetes mellitus") ("sensory neuropathy") (treatment OR therapy)
- Why important: simple way to ensure the search engine evaluate your query the way you want, from left to right



- Example: ("sensory neuropathy") AND ("diabetes mellitus") AND (biguanide\*)
   AND (treatment OR therapy)
- Why important. AND glues the query together



### Put Your Main Concept First

- Example: ("sensory neuropathy") ("diabetes mellitus") (treatment OR therapy)
- Why important: put main subject first. Engines tend to rank documents more highly that match first terms or phrases evaluated

### Avoid Misspellings

**searching** 269,000,000

تنايات التنشادات

- **serching** 207,000
- searchng 97,700
- **seerching** 3,860
- **Sherching** 5,670



Martindale Center

http://www.martindalecenter.com/

Hardin Website

http://www.lib.uiowa.edu/hardin/md/

## Academic Search Engines

- Google Scholar
- http://scholar.google.com



- Scirus
- http://www.scirus.com







### Google Scholar

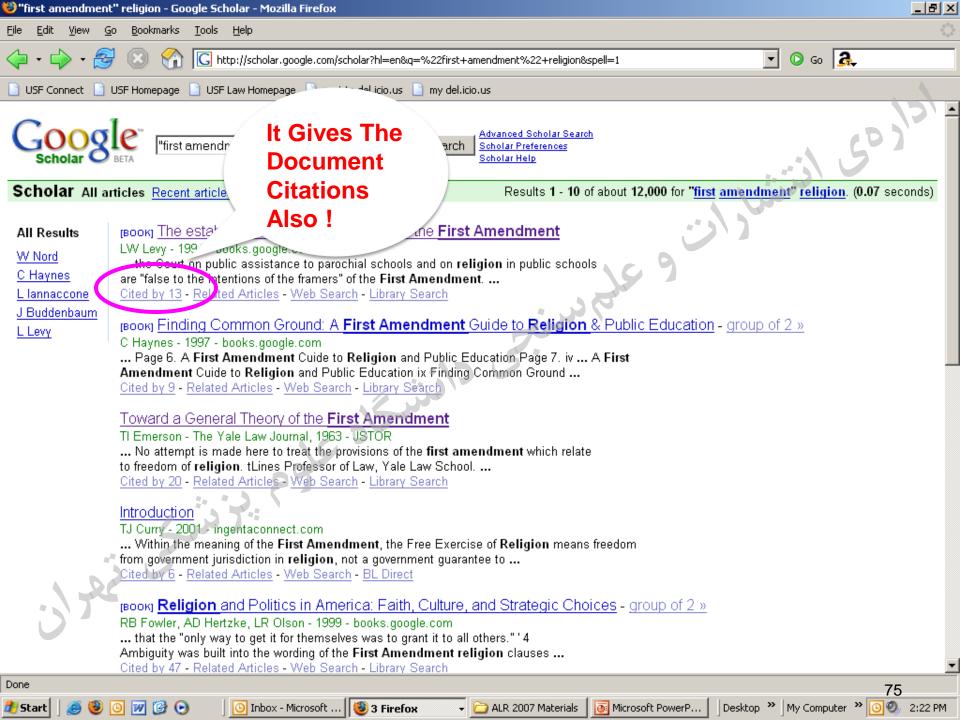
Google Scholar provides a simple way to broadly search for scholarly literature including: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites.



- Covers: medicine, law, social sciences, arts, humanities, business, & finance.
- Included items: peer-reviewed papers, theses, book excerpts, abstracts & full-text articles
- Sources for items: academic publisher web pages, professional societies, preprint repositories, universities, & other scholarly organizations.

## Features of Google Scholar

- Search all scholarly literature from one convenient place
- Explore related works, citations, authors, and publications
- Locate the complete document through your library or on the web
- Keep Authors Profiles
- Check who's citing your publications, create a public author profile



## **Authors Profiles Services**

- Through these services, you can set your own academic CV's & profiles & make them visible in the web for all.
- 1) ResearcherID
- 2) Scopus Authors ID
- 3) ORCID (Open Researcher & Contributor ID)
- 4) Google Citation Service

# Google Scholar Citation Service



Sign up for a new Google Account

#### Citations

Welcome to Google Scholar Citations

Track citations to your publications

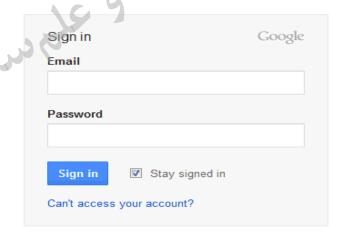
Check who is citing your publications. Graph your citations over time. Compute citation metrics.

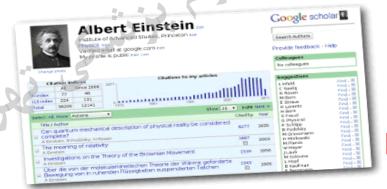
View publications by colleagues

Keep up with their work. See their citation metrics

Appear in Google Scholar search results

Create a public profile that can appear in Google Scholar when someone searches for your name.





http://scholar.google.com/citations



Step 2: Articles Step 3: Updates



#### **Profile**

Track citations to your publications. Appear in Google Scholar search results for your name.

Name:\*

Bagher Larijani

Affiliation:

For example: Professor of Computer Science, Stanford University

Tehran University of Medical Sciences,

Email for verification:

Use an email address at your institution. For example: yourname@mit.edu

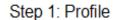
larijani@tums.ac.ir

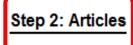
Areas of interest:

For example: Artificial Intelligence, Conservation Biology, Pricing Theory

Endocrinology, Medical Ethics

Next step





Step 3: Updates



#### Add articles - Bagher Larijani

Find articles that you've written and add them to your profile. Later, you can edit or delete the articles in your profile or add more articles to your profile.

author: "Bagher Larijani"

Search article groups

Try searching for your name, article titles, co-authors, or topical keywords.

#### Article groups

#### Bagher Larijani

A review on the role of antioxidants in the management of diabetes and its complications

R Rahimi, S Nikfar, B Larijani... - Biomedecine & Pharmacotherapy, 2005

Vitamin D deficiency and causative factors in the population of Tehran

S Hashemipour, B Larijani, H Adibi, E Javadi... - BMC Public health, 2004

Add all 598 articles | See all articles

author: "Bagher Larijani"

Search article groups

1.1

Step 1: Profile

Step 2: Articles

Step 3: Updates



Added article group.

#### Add articles - Bagher Larijani

Articles: 598 Citations: 3279

Next step

Find articles that you've written and add them to your profile. Later, you can edit or delete the articles in your profile or add more articles to your profile.

author: "Bagher Larijani"

Search article groups

Try searching for your name, article titles, co-authors, or topical keywords.

#### Article groups

1-1

#### Bagher Larijani

A review on the role of antioxidants in the management of diabetes and its complications

R Rahimi, S Nikfar, B Larijani... Biomedecine & Pharmacotherapy, 2005

Vitamin D deficiency and causative factors in the population of Tehran

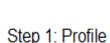
S Hashemipour, B Larijani, H Adibi, E Javadi... - BMC Public health, 2004

Remove all 598 articles | See all articles (All articles are already in your profile)

author: "Bagher Larijani"

Search article groups

http://scholar.google.com/citations



Step 2: Articles

Step 3: Updates



#### Profile updates

- We'll use a statistical authorship model to identify new articles that you write. We may also update bibliographic information for articles in your profile or identify duplicate article entries, which could be merged or deleted. How would you like to handle these changes?
  - Automatically update the list of articles in my profile. (recommended)
  - Don't automatically update my profile. Send me email to review and confirm updates.
- You can also add and remove individual articles, update their bibliographic data, and merge duplicate records. Rest assured, our automatic updates will preserve your edits and will not override them.
- We'll collect and display citations to your articles from all of Google Scholar. The citations will update automatically to reflect changes in your profile and in Google Scholar.

Go to my profile

http://scholar.google.com/citations

Your profile is private and won't appear in search results. (Make my profile public - Preview public version)

Citations to my articles



#### Bagher Larijani Edit

Tehran University of Medical Sciences Edit Endocrinology - Medical Ethics Edit Email at tums.ac.ir Edit (pending verification Resend) My profile is private Edit Add homepage

Change photo

Citation indices

multicenter study in Tehran

BMC Endocrine Disorders 3 (1), 4

Since 2006

		/ 111	Office 2000										
Citations		3279	3119										
h-i	index	27	26			_							
i10	0-index	104	101	0-	2004	2005	10	2007		2009			2011
Select: All, None Actions ▼				•						Show:	20 🕶	1-20	Next >
	Title / Au	thor			4	10					Cit	ed by	Year
	A review on the role of antioxidants in the management of diabetes and its complications R Rahimi, S Nikfar, B Larijani, M Abdollahi Biomedecine & Pharmacotherapy 59 (7), 365-373										188	2005	
	Vitamin D deficiency and causative factors in the population of Tehran  S Hashemipour, B Larijani, H Adibi, E Javadi, M Sedaghat, M Pajouhi, A  BMC Public health 4 (1), 38									98	2004		

Metabolic and endocrinologic complications in beta-thalassemia major: a

A Shamshirsaz, MR Bekheirnia, M Kamgar, N Pourzahedgilani, N Bouzari, M ...

### Google scholar

Search Authors

My Citations - Help

lext >

2003



Suggested co-authors									
R Heshmat	Find - 🗵								
A Hossein-nezhad	Find - 🗵								
H Fakhrzadeh	Find - 🗵								
F Zahedi	Find - 🗵								
M Pajouhi	Find - 🗵								
A Soltani	Find - 🗵								
M Ganjali	Find - 🗵								
P Norouzi	Find - 🗵								
V Haghpanah	Find - 🗵								
M Abdollahi	Find - 🗵								
Z Maghbooli	Find - 🗵								
M Amoli	Find - 🗵								
F Bandarian	Find - 🗵								
S Tavangar	Find - 🗵								
F Faridbod	Find - 🗵								

# الانتاع الانتناد الدن Internet is like a library **But a Library** with many problems

- The Web lacks the bibliographic control standards we take for granted in the print world;
- There is no equivalent to the ISBN to uniquely identify a document
- There is no standard system, analogous to those developed by the library of congress, of cataloguing or classification
- There is no central catalogue including the Web's holdings; in fact, many, if not most, Web documents lack even the name of the author and the date of publication.
- There is no content control.

# \_anguage Bias Database Bias Publication Picture Com **Avoid Bias in Information**

- Commercial Bias
- Personal opinion Bias



#### Health On the Net Foundation

Non Governmental Organization

Medical information you can trust!



| SPEECH: ON - More info? | FR - EN - DE - SP - CN - PL - NL | HOME



PATIENT / INDIVIDUAL

#### MEDICAL PROFESSIONAL

WEB PUBLISHER

#### The services offered by HON









HONcode

HONcode certification: Improving the quality of online health information

#### **HONsearch**

Search only reliable and trustworthy medical websites

**HONtools** 

Use our free services to enhance and improve your online experience

**HONtopics** 

Access a varied list of reliable medical/health topics



- Created in 1995, The Health On the Net Foundation (HON) is a non-profit, nongovernmental organization, accredited to the Economic and Social Council of the United Nations.
- HON promotes and guides the deployment of useful and reliable online health information.

■ For 15 years, HON has focused on the essential question of the provision of health information to citizens, information that respects ethical standards.