

CS4051

Information Retrieval

Week 01

Muhammad Rafi

January 22, 2024

Agenda

- Course Introduction
 - Basic Terminology
 - Structured Vs. Semi Structured Vs. Unstructured
 - Difference between Database & IR
 - Ad hoc IR setup
 - Conclusion
-

Course Objectives

- Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources usually document collections.
- The course discusses important retrieval models (Boolean, vector space, probabilistic, inference net, language modeling, link analysis), clustering algorithms, collaborative filtering, automatic text categorization, and experimental evaluation.
- Search Engines Algorithms, Advanced IR techniques and Evaluations are discussed.

Course Outline (1/4)

- Week -01
 - Introduction to IR course, IR Problem and its Components, Basic IR model – Boolean Information Retrieval, Extended Boolean Model, Example of Commercial Systems – WestLaw <Ch: 1 >
- Week-02
 - Document Processing, Term vocabulary and positing lists, Stemming & Lemmatization, posting list processing via skip lists, Phrase Query, positional indexing, bi-word indexing, Combining different indexing techniques. <Ch: 2>

Course Outline (1/4)

■ Week -03

- Dictionary and Tolerant Retrieval, Search Structures for Dictionary, Wildcard queries, permuterm index, k-gram index, Spelling Correction, Edit Distance, Phonetic Correction <Ch: 3>

■ Week-04

- Index construction, single pass, in-memory, distributed indexing, dynamic indexing.
- Heaps law, Zipf's law, dictionary indexing, fixed length and variable length coding, γ codes <Ch:4 and 5>

Course Outline (2/4)

■ Week -05

- Vector Space Model <Ch: 6 and Ch:7>

■ Week-06

- Midterm Exam

■ Week-07

- Evaluation of IR <Ch: 8>

■ Week-08

- Relevance Feedback <Ch: 9>

Course Outline (3/4)

- Week-9
 - Basic Web Search, Crawler and Indexing <Ch: 19 and 20>
 - Week-10
 - Link Analysis <Ch: 21>
 - Week-11
 - Midterm II
 - Week-12
 - Text classification <Ch: 14>
-

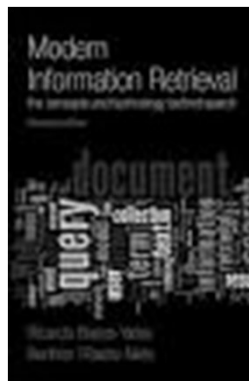
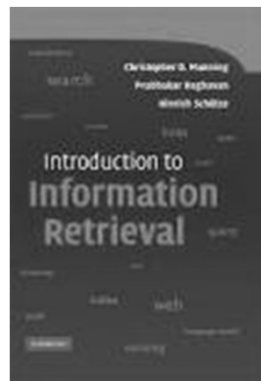
Course Outline (4/4)

- Week-13
 - Text Clustering
 - Week-14
 - Question/ Answering
 - Week-15
 - Interactive Information Retrieval
 - Week-16
 - Neural Information Retrieval
-

Grading Scheme

■ Assignments	20%
■ Quizzes	10%
■ Mid-Terms (2)	20%
■ Class Project	10%
■ Final	40%

Text Books/References



Information Retrieval

- Information Retrieval (IR) is activity of finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need (query) from within large digital collections (usually stored on computers).
- Example
 - What is your CNIC Number? CC?
 - What is De Quervain's?
 - How to get rid from kitchen bugs?
 - What is Surakav?
 - What is Sarcopenia?

Example

Google X 🔊 🔍


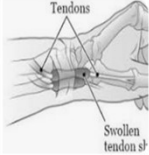
[Q](#) All [Images](#) [Videos](#) [Books](#) [News](#) [More](#) [Tools](#)

About 2,000,000 results (0.53 seconds)

De Quervain's tenosynovitis is a **painful condition that affects the tendons in your wrist**. It occurs when the 2 tendons around the base of your thumb become swollen. The swelling causes the sheaths (casings) covering the tendons to become inflamed. This puts pressure on nearby nerves, causing pain and numbness. 07-May-2020

<https://familydoctor.org/de-quervains-tenosynovitis/> ⋮

de Quervain's Tenosynovitis: Causes & Treatment

About featured snippets • Feedback

People also ask ⋮

How long does it take for de Quervain to heal? ▼

Can De Quervain's be cured? ▼

Terminology

- Data / Information / Knowledge/ Wisdom
- Database systems Vs. Information retrieval
- Data Mining Vs. Text mining
- Text Mining Vs. Information retrieval
- Ad hoc Information Retrieval

GCR – Code



Database Systems

- Generally data collected by some database management systems are collectively called Database.
 - Mostly structured one, to support business operations.
 - SQL is used to get the required information from it. Routines Vs. Ad hoc queries.
-

IR Vs. Databases

- Format of data:
 - DB: Structured data. Clear semantics based on a formal model.
 - IR: Mostly Semi-Structured or unstructured. Free text.
 - Queries:
 - DB: Formal (like SQL)
 - IR: often expressed in natural language (keywords search)
 - Result:
 - DB: exact result
 - IR: Sometimes relevant, often not
-

Data Mining

- Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD). It is an interdisciplinary subfield of computer science
 - The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.
-

Text Mining

- "The objective of Text Mining is to exploit information contained in textual documents in various ways, including ...discovery of patterns and trends in data, associations among entities, predictive rules, etc." (Grobelnik et al., 2001)
 - "Another way to view text data mining is as a process of exploratory data analysis that leads to heretofore unknown information, or to answers for questions for which the answer is not currently known." (Hearst, 1999)
-

Text Vs. Data

	Data Mining	Text Mining
Data Object	Numerical & categorical data	Textual data
Data structure	Structured	Unstructured & semi-structure
Data representation	Straightforward	Complex
Space dimension	<tens of thousands	> tens of thousands
Methods	Data analysis, machine learning statistic, neural networks	Data mining, information retrieval, NLP,...
Maturity	Broad implementation since 1994	Broad implementation starting 2000
Market	10 ⁵ analysts at large and mid size companies	10 ⁸ analysts corporate workers and individual users

Information Retrieval Vs. Text Mining

- Information Retrieval is more concerned with getting information related to a user query from unstructured collections of generally textual data. Activity like:
 - Crawling, parsing, indexing and evaluating the retrieved information by using generalized methods
- Text Mining generally utilizes the NLP based methods for doing activity like:
 - Clustering, classification, summarization and QA

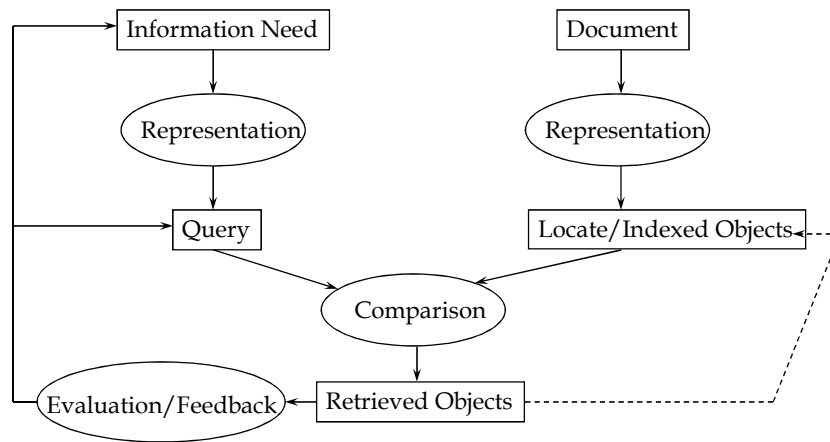
Information Retrieval Vs. Information Extraction

- Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources.
 - Searches can be based on metadata or on full-text indexing.
 - Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents.
-

Adhoc Retrieval Systems

- It is a system that aims to provide documents from within the collection that are relevant to an arbitrary user information need, communicated to the system by means of a one-off, user-initiated query.
 - Example
 - Google
 - Bing (later Decision engine)
-

Basic Information Retrieval Process



Content wise IR Systems

■ Contents

- Size
 - Personal / Desktop (Spotlight, Instant Search, anywhere)
 - Web Scale (Google, DuckDuckgo)
 - Enterprise Search (WestLaw, PolicyBazar)
- Static (Offline) or Dynamic (Online)
- Type
 - Text
 - Multimedia
 - Mixed
- Exact Match vs. Best Match

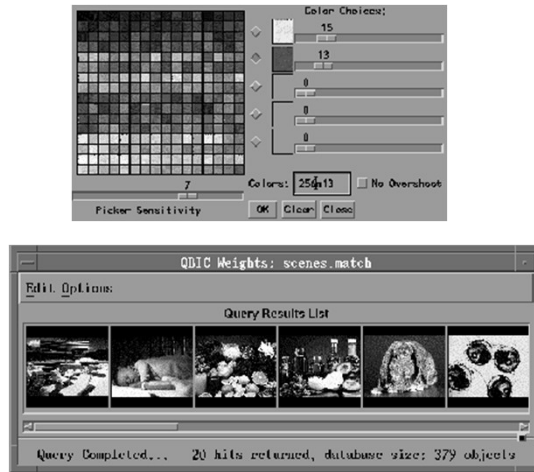
Dimension of IR

Content	Applications	Tasks
Text	Web search	Ad hoc search
Images	Vertical search	Filtering
Video	Enterprise search	Classification
Scanned docs	Desktop search	Question answering
Audio	Forum search	Clustering
Music	P2P search	
	Literature search	

IBM's QBIC

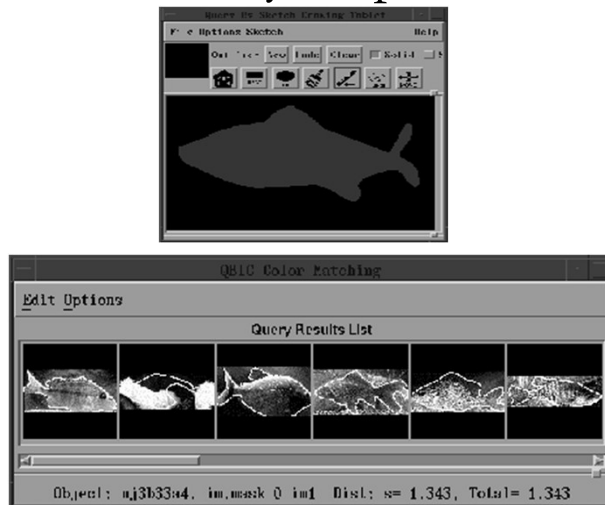
- QBIC – Query by Image Content
- First commercial CBIR system.
- Model system – influenced many others.
- Uses color, texture, shape features
- Text-based search can also be combined.
- Uses R*-trees for indexing

QBIC – Search by color



**** Images courtesy : Yong Rao**

QBIC – Search by shape



**** Images courtesy : Yong Rao**

QBIC – Query by sketch



**** Images courtesy : Yong Rao**

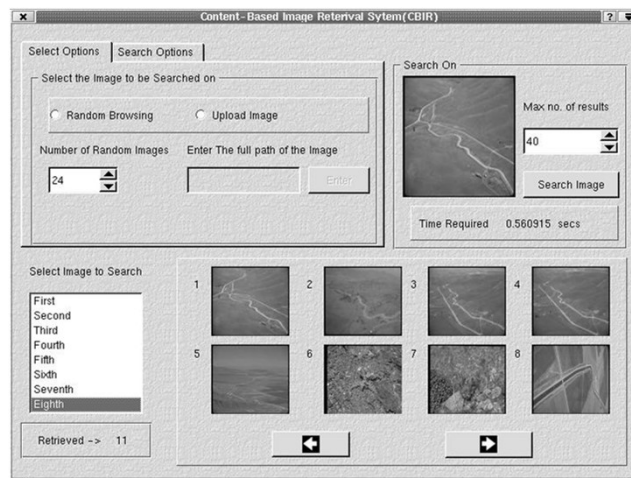
Virage

- Developed by Virage inc.
- Like QBIC, supports queries based on color, layout, texture
- Supports arbitrary combinations of these features with weights attached to each
- This gives users more control over the search process

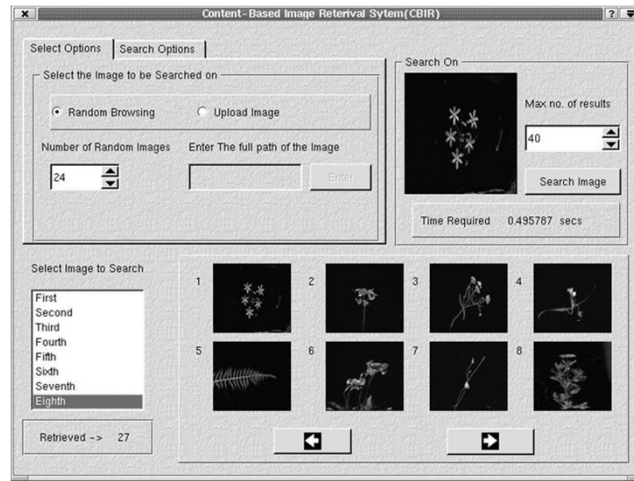
VisualSEEk

- Research prototype – University of Columbia
- Mainly different because it considers spatial relationships between objects.
- Global features like mean color, color histogram can give many false positives
- Matching spatial relationships between objects and visual features together result in a powerful search.

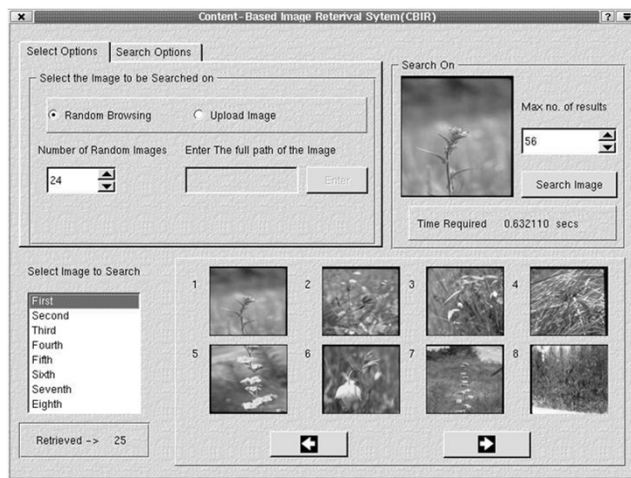
ISearch



ISearch



ISearch



Example: Content-based Image Retrieval:

□ <http://wang.ist.psu.edu/IMAGE>



Music Search Engine

- Search by metadata likes: artists biography, music reviews, new releases, concerts dates
- Search for Lyrics: Lyrics.com. Allmusic.com musicpedia.com and SearchLyrics.com
- Recommend Similar Music: based on seed elements (artist, track) users are recommended similar tracks or artists.
- User query by Humming
- Generate Playlists: automatic generation of playlists, that satisfy user constraints

Mathematical Search Engines

- The mathematical contents on the web is continuously on rise. The IR perspective of contents are quite low.
 - We need to redesign a specialized search engine for it.
 - There are few very good attempts:
 - www.wolframalpha.com
 - symbolab.com
 - searchenginewatch.com
-

Question & Answering Systems

- System that enables the extraction of an answer (or several) to a request (a question) based on a corpus
 - An answer or several, possibly a list from one or several documents, an answer of the type Yes/No...,
 - Askjevees.com
 - Easyask.com
 - Answerlogic.com
-

General Search Engines

- www.google.com
 - www.visimo.com
 - www.clusty.com
 - www.bing.com
 - www.yahoo.com
-

Bing

- MSN Search
 - Microsoft Search
 - Live Search
 - Bing
 - ASP.NET Launched June 01, 2009 supported 40 languages now
 - Bing and Decide was the slogan in 2009
 - Bing a decision engine - 2010
 - “Bing is for doing” is in 2012
-

YAHOO! HOMEPAGE (1994)

Yahoo - A Guide to WWW

[What's New? | What's Cool? | What's Popular? | Stats | A Random Link]

[Y Top](#) [Up](#) [Search](#) [Mail](#) [Add](#) [Help](#)


- [Art\(466\)](#) NEW
- [Business\(6426\)](#) NEW
- [Computers\(2609\)](#) NEW
- [Economy\(743\)](#) NEW
- [Education\(1487\)](#) NEW
- [Entertainment\(6199\)](#) NEW
- [Environment and Nature\(193\)](#) NEW
- [Events\(53\)](#) NEW
- [Government\(1031\)](#) NEW
- [Health\(367\)](#) NEW
- [Humanities\(163\)](#) NEW
- [Law\(163\)](#) NEW
- [News\(185\)](#) NEW
- [Politics\(148\)](#) NEW
- [Reference\(474\)](#) NEW
- [Regional Information\(2606\)](#) NEW
- [Science\(2634\)](#) NEW
- [Social Science\(93\)](#) NEW
- [Society and Culture\(648\)](#) NEW

<--! Removed by Correct v2.0 by Kutsal Berberoglu --> 23836 entries in Yahoo [[Yahoo](#) | [Up](#) | [Search](#) | [Mail](#) | [Add](#) | [Help](#)]

yahoo@akebono.stanford.edu

Copyright © 1994 David Filo and Jerry Yang

YAHOO! HOMEPAGE (1995)



•NEW• COOL•RANDOM•

HEAD YAHOO ADD
LINES INFO URL

Now Open: [Yahoo! Surf Shop!](#) [Remove the Snapple Gap to Play](#) [Web Launch](#)

[Options](#)

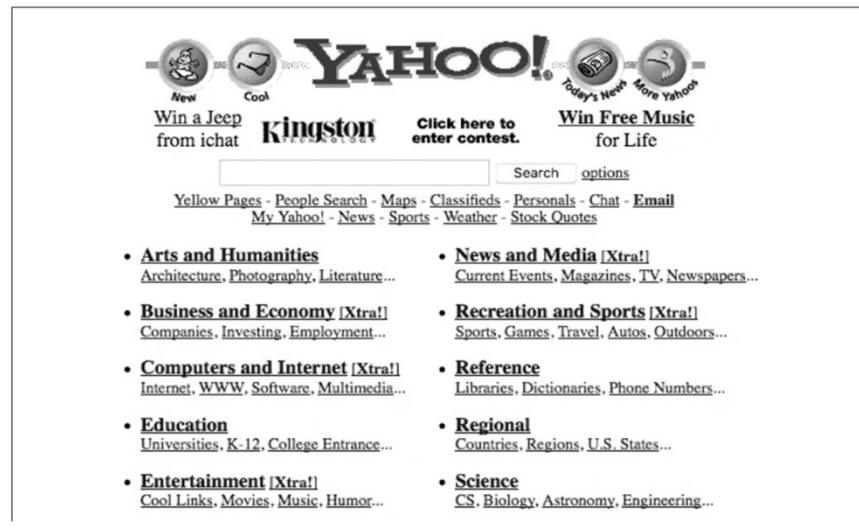
- **Arts**
Humanities, Photography, Architecture, ...
- **Business and Economy [Xtra!]**
Directory, Investments, Classifieds, Taxes, ...
- **Computers and Internet [Xtra!]**
Internet, WWW, Software, Multimedia, ...
- **Education**
Universities, K-12, Courses, ...

- **News [Xtra!]**
World [Xtra!], Daily, Current Events, ...
- **Recreation**
Sports [Xtra!], Games, Travel, Autos, ...
- **Reference**
Libraries, Dictionaries, Phone Numbers, ...
- **Regional**
Countries, Regions, U.S. States, ...

YAHOO! HOMEPAGE (1996)



PERSONALIZED YAHOO! (1997)



YAHOO! HOMEPAGE (1998)



YAHOO!

What's New Check Email Personalize Help

World Cup 98 **SURPLUS DIRECT** 3 speed Transmission Yahoo! Mail
free @yahoo.com email

Search options

Yahoo! Games - play online chess, backgammon, bridge, blackjack and more...

Yellow Pages - White Pages / People Search - Maps - Classifieds - Personals - Chat - Email
Shopping - My Yahoo! - News - Sports - Weather - Stock Quotes - more...

- Arts and Humanities**
Architecture, Photography, Literature...
- Business and Economy [Xtra!]**
Companies, Finance, Employment...
- Computers and Internet [Xtra!]**
Internet, WWW, Software, Multimedia...
- Education**
Universities, K-12, College Entrance...
- News and Media [Xtra!]**
Current Events, Magazines, TV, Newspapers...
- Recreation and Sports [Xtra!]**
Sports, Games, Travel, Autos, Outdoors...
- Reference**
Libraries, Dictionaries, Phone Numbers...
- Regional**
Countries, Regions, U.S. States...

Y2K YAHOO! HOMEPAGE (2000)



YAHOO!

Auctions Messenger Check Email What's New Personalize Help

Yahoo! Auctions coins, cards, stamps, cars Know when friends are online! Click to download Yahoo! Messenger Yahoo! Mail free email for life

Search advanced search

Shop: Auctions: Classifieds: Shopping: Travel: Yellow Pgs: Maps: Media: News: Sports: Stock Quotes: TV: Weather: Connect: Chat: Clubs: Games: GeoCities: Greetings: Invites: Mail: Messenger: Personals: People Search: For Kids: Personal: My Yahoo!: AddrBook: Calendar: Briefcase: Photos: Alerts: Bookmarks: Companion: Bill Pay: more...

Yahoo! Shopping - Thousands of stores. Millions of products.

Departments	Stores	Features
Apparel	Gap	J.K. Rowling
Flowers	I-800-Flowers	Gift Ideas
Luxury	Banana Republic	Special Offers
Computers	Macy's	Digital Cameras
Music		
Electronics		
Video/DVD		

Arts & Humanities
Literature, Photography...

Business & Economy
B2B, Finance, Shopping, Jobs...

Computers & Internet
Internet, WWW, Software, Games...

News & Media
Full Coverage, Newspapers, TV...

Recreation & Sports
Sports, Travel, Autos, Outdoors...

Reference
Libraries, Dictionaries, Quotations...

In the News
DNS official convicted of espionage
New US dietary guidelines
Upstart Transmeta's chip tapped for AOL devices
Playoffs - NBA, NHL
more...

Marketplace
Free 56K Internet Access
Y! Travel - buy tickets, check arrival times
Looking for a car? job? house? date?

Broadcast Events
2:20pm ET: Braves vs. Cubs

YAHOO! HOMEPAGE (2005)

Yahoo! Search

Yahoo! Messenger - Hide from Clyde. Be seen by Gene. Get control with Stealth settings.

Free mail Sign Up

Autos
Chat
Finance
Games
GeoCities
Groups
Health

Horoscopes
HoJobs
Kids
Maps
Messenger
Mobile

Movies
Music
My Yahoo!
News
People Search
Personals
Photos

Real Estate
Shopping
Sports
Travel
TV
Yellow Pages
All Y! Services...

ONLINE DEGREES
• Associate • Bachelor's
• Master's • Postgrad
START TODAY!
classesUSA

In the News

- Report: Earthquake in Iran kills 175
- Bush seeks more NATO support on Iraq
- Palestinians face crisis over new Cabinet
- N. Korea indicates return to nuclear talks
- Group honors Malcolm X at NYC theater
- Mexico schools' songbook spurs drug debate
- Bill to ban cat, dog eating fires ethnic groups
- NBA - Daytona 500 - NCAA Hoops - Tennis

News - Popular - Sports - Stocks

Weather
Enter City or U.S. Zip Code
Go

YAHOO! HOMEPAGE (2013)

Web Search

Sign In Mail

Star transformed to play Anna Nicole Smith
The actress in the role of the late Playboy model dyed her dark locks platinum, wore prosthetics, and changed her accent. See the whole cast.

Selena Gomez Goes Without Makeup, Shows Lighter Hair Color
The "Come & Get It" singer was photographed with a fresh face and ombre locks in London on May 21. Us Weekly

Column: Garcia will pay dearly for Tiger remarks
Sergio Garcia could have stopped after he whined about fans of Tiger Woods making too much noise on the golf course. He should have stopped after saying Woods is not a nice person who has been lying to the media since the first time he The Associated Press

Trending Now

- Wayne Miller dies
- Janet Jackson billionaire
- Gosling film booed
- Zach Sobiech dies
- Tesla \$465 million
- Sisters ejected from ...
- Drone strikes 4 Amer...
- Bieber may sue guests
- Pauly D new hair
- Free burgers for life

THE TOP HEADLINES. THE MOST INTERESTING CHIT-CHAT.
Get all the latest on yahoo.com.

San Francisco
50°F Partly Cloudy

Today 60°-69°
Tomorrow 59°-69°
Friday 61°-71°

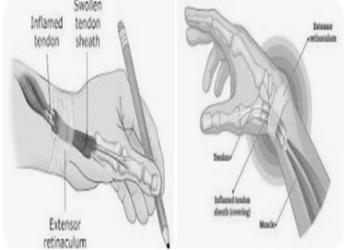
Google What is De Quervain's? X ↻ 🔍

Images Videos Thyroiditis Tenosynovitis Release Shopping News Books Finance All filters Tools SafeSearch


About 1,030,000 results (0.42 seconds)

🔊 اردو میں 🗣️ In English


De Quervain's tenosynovitis is painful swelling around your thumb tendons. Doing repetitive motions with your thumbs and wrist at work or for a hobby are the most common causes. Most people need to wear a splint and rest their wrist for a few weeks to recover. 14-Jun-2023



De Quervain's tenosynovitis is painful swelling around your thumb tendons. Doing repetitive motions with your thumbs and wrist at work or for a hobby are the most common causes. Most people need to wear a splint and rest their wrist for a few weeks to recover. 14-Jun-2023

 Cleveland Clinic
<https://my.clevelandclinic.org/health/diseases/1091...>

De Quervain's Tenosynovitis: Symptoms and Treatment



🔍 About featured snippets • 🗉 Feedback

People also ask :

- What causes de Quervain's syndrome? ▾
- How do you fix de Quervain syndrome? ▾
- What is de Quervain's also known as? ▾

**You**

what is De Quervain's

**ChatGPT**

De Quervain's tenosynovitis, also known as De Quervain's syndrome or De Quervain's disease, is a condition that affects the tendons on the thumb side of the wrist. The condition is characterized by inflammation of the tendons that control the movement of the thumb. The specific tendons involved are the abductor pollicis longus and extensor pollicis brevis.

The inflammation can cause pain and swelling near the base of the thumb, making it difficult to move the thumb and grip objects. Repetitive hand and wrist movements, as well as activities that involve forceful gripping or pinching, are common causes of De Quervain's tenosynovitis.

Typically, the symptoms include pain along the back of the thumb and wrist, especially when gripping or rotating the wrist. Swelling and tenderness may also be present.

Treatment options for De Quervain's tenosynovitis may include rest, immobilization of the thumb and wrist, anti-inflammatory medications, and physical therapy. In more severe cases, corticosteroid injections or surgical intervention may be considered.