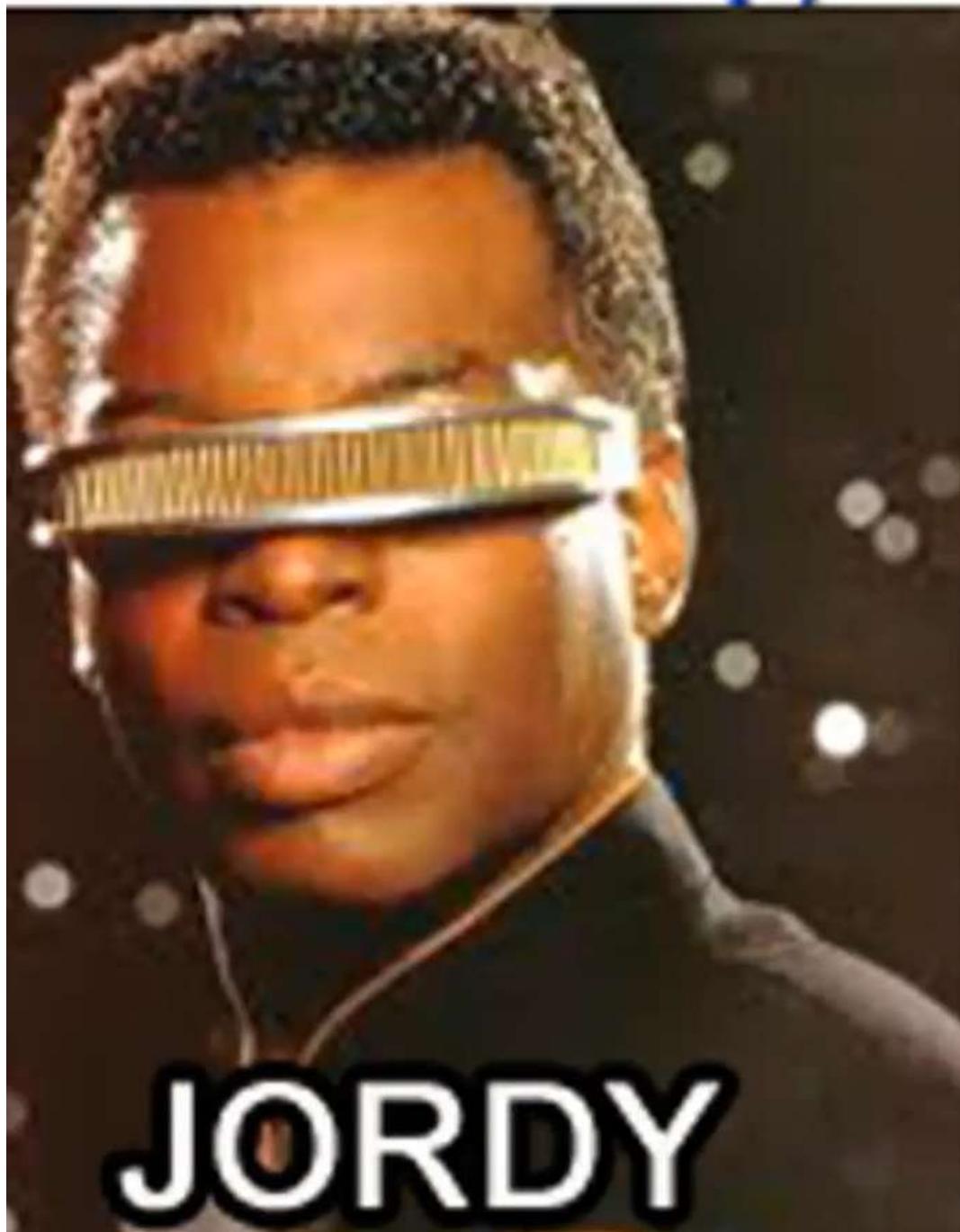


# Other example (talking about view...!)



# Google GLASS



**JORDY**



**JORDY 2.0**

10:42

"ok glass"

ok glass,google...  
Ask a question



Web results  
for  
awesome

awesome

now





# GOOGLE GLASS

EVERY TIME I SEE SOMEONE WITH A  
GOOGLE GLASSES , I'M GOING TO GO  
UP TO THEM AND SCREAM "GOOGLE  
GLASSES : IMAGE SEARCH HORSE  
FUCK. SAFE SEARCH OFF. OPEN FIRST  
50 RESULTS IN NEW TABS...

i WILL THEN RUN OFF INTO THE NIGHT..



# Another essential feature...

- ◆ ... For the success of a site, is to be visible ***outside***
- ◆ Typically, via ***search engines***

# So, it is critical...

- ◆ To be ranked highly from a search engine
- ◆ → in general, in the **SERP**  
**(Search Engine Results Page)**

# And so, to understand search engines

- ◆ And to understand how they assign points to a page...

# Besides, of course...

- ◆ Even more important, to understand how users behave (!)

# For instance...

- ◆ How much is it important to appear in the top ten?
- ◆ And, within the top ten, how much the ranking is important??

# Top ten or not top ten?

- ◆ La top ten absorbs...
- ◆ ... more than **95%** of all clicks!!



# First position?

- ◆ 1) **51%** of clicks (!!!!!!!)
- ◆ So the first positions already attracts  
***more clicks than all the others combined!!***



# Go on...

- ◆ 2) 16% of clicks (!)
- ◆ 3) 6%
- ◆ 4) 6%
- ◆ 5) 5%
- ◆ 6) 4%
- ◆ 7) 2%
- ◆ 8) 1%
- ◆ 9) 1%



# The “black jersey” effect

- ◆ Also known as “Malabrocca effect”

# Luigi Malabrocca



So: “Malabrocca effect”





- Malabrocca, born in Tortona on June 22, 1920, also called "the Chinese" because of his almond-shaped eyes, was a runner's finest. In his career he has won 138 races, 15 of which as a professional (Milan-Sanremo 1948, Tour of Croatia and Slovenia 1949), and was twice Italian cyclocross champion (1951 and 1953).

He was running away behind the group. He went into bars and never went out. He hid in the escarpments, in the barns, in the cellars. Once he even dived into a kind of well, empty, but a mustachioed farmer, his skin browned by the sun, became suspicious and lifted the lid: "So what?", He ordered him. "I'm racing the Giro d'Italia ", he tried to explain. Then he got back on the bike, faced Rolle, Pordoi, Campolongo and Gardena, because it was the dolomitic stage of the Giro d'Italia, and finally reached the finish line. Last, very last, black

the barrels, in the cellars. Once he even dived into a kind of well, empty, but a mustachioed farmer, his skin browned by the sun, became suspicious and lifted the lid: "So what?", He ordered him. "I'm racing the Giro d' Italia ", he tried to explain. Then he got back on the bike, faced Rolle, Pordoi, Campolongo and Gardena, because it was the dolomitic stage of the Giro d'Italia, and finally reached the finish line. Last, very last, black jersey, very black. It was his strong point, his very strong one.

Corsi | EasyBadge

/D/MM/UNIPD/WIM-partial/lezioni/lezione9/malab-EN.shtml

400% Search

file:///D/MM/UNIPD/WIM-partial/lezioni/lezione9/malab-EN.shtml

Last in 1946 at 4.9'34" by Gino Bartali: a feat. Last in 1947 at 5.52'20" by Fausto Coppi: a masterpiece. One man alone in command, Malabrocca, but from the bottom of the standings, until he met another phenomenon in the fight against the deadline, Sante Carollo. It was the 1949 Giro, Carollo had two hours of advantage and the final stage, Turin-Monza, with official arrival in Milan, did not propose ambushes. "Il Mala" thought about it: while Carollo was pedaling unaware in the middle of the group he took advantage

"Mala" thought about it: while Carollo was pedaling unaware in the middle of the group, he took advantage of a puncture, entered a tavern, first accepted a drink, then an invitation to the home of a fan who wanted to show him a particular equipment for fishing, finally he got back on the saddle and pedaled at idle. A triumph in reverse: two hours and 20 behind the winner Giovannino Corrieri, two hours and a quarter behind Carollo. But Luisin had made a mistake: he had not foreseen that the timekeepers - once so impatient - had

reverse: two hours and 20 behind the winner Giovannino Corrieri, two hours and a quarter behind Carollo. But Luisin had made a mistake: he had not foreseen that the timekeepers - once so impatient - had already returned home, classifying the super late with the same gap as the group. So it was Carollo who won the black jersey: 9.57'07" by Coppi against 7.47'26" collected by Malabrocca. And Luisin, disappointed, took the solemn decision to abandon that divine human comedy.

## So: “Malabrocca effect”

- ◆ At the last place in the top teb, the click-rate ***doubles*** compared to the previous positions: **2% !!**





spanish water dog

Search

Advanced Search  
Preferences

Web Images

Results 1 - 10 of about 1

### Spanish Water Dog Advisor Line American Spanish Water Dog

Care of dogs, adoption applications for new and dogs, photographs, history, pedigree, training, health, and more.

www.spanishwaterdog.org

### Spanish Water Dog | Spanish Water Dogs Spanish Water Dog Water Dogs

about the Spanish Water Dog, its peculiar features, care, temperament, health problems, pregnancy and moulting.

www.spanishwaterdog.org/spanishwaterdog.htm#Spain

### U.S.A. Spanish Water Dog

do Chesapeake Kennel dedicated to preserving the original qualities of this breed joined in North America.

The Spanish Water Dog is a breed of dog developed by the

Spaniards in Spain. Spanish Water Dog who were also used in hunting of deer developed by the

Spanish Water Dog. Spanish Water Dog is a breed of dog developed by the

### Spanish Water Dog Club, Inc.

Spanish Water Dog Club (SWD Club), an ABC-based club dedicated to the health and well-being of the Spanish Water Dog (Perro de Agua Espanol).

www.spanishwaterdogclub.org

### De Fabio Spanish Water Dogs

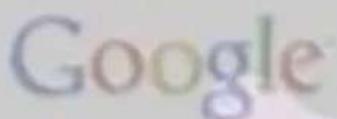
Photographs, a buyer's guide, breed standards, and articles about the breed. American Fox.

www.spanishwaterdogs.com/americanfox.htm

### American Kromfohrländer Spanish Water Dog

The Spanish Water Dog is a ancient and noble breed. It was first mentioned in historical

documents in 1110 AD. It is also known as Spanish Water Dog.



spanish water dog

Search

Advanced Search  
Preferences

Web Images

Results 1 - 10 of about 1

[Spanish Water Dog Association of America](#)  
Coches de agua - a traditional Spanish breed of dog. This page has links to the breed's history and other web sites.

**Spanish Water Dog** - Wikipedia, the free encyclopedia. Spanish Water Dog. The Spanish Water Dog, also known as the Spanish Water Spaniel, is a breed of dog developed in Spain. It is a small, long-legged, long-bodied dog with a thick, wavy coat.

**U.S.A. Spanish Water Dogs**  
de Chaves - a kennel dedicated to the long-term breeding of this great breed in North America. The Spanish Water Dog is a breed of dog developed in Spain. It is a small, long-legged, long-bodied dog with a thick, wavy coat.

**Spanish Water Dog - Wikipedia, the free encyclopedia**  
The Spanish Water Dog (Spanish: "Perro de Agua Español") is a breed of dog developed by the shepherds in Spain. Developed probably by one who was also used to another type of dog developed by the Moors in Spain, the Spanish Water Dog is a very similar type of dog. It is a small, long-legged, long-bodied dog with a thick, wavy coat.

**Spanish Water Dog Club Inc.**  
Spanish Water Dog Club (SWD Club), an AKC-based club dedicated to the long-term health and welfare of Spanish Water Dogs (Perro de Agua Espanol). We are a club dedicated to the long-term health and welfare of Spanish Water Dogs (Perro de Agua Espanol).

**De Pinto Spanish Water Dogs**  
Photographs, a buyer's guide, breed standards, and articles about the breed. American Fox, Utah, USA - De Pinto Spanish Water Dogs. De Pinto Spanish Water Dogs are a breed of dog originally from Spain. They are a small, long-legged, long-bodied dog with a thick, wavy coat.

**American Kennel Club: Spanish Water Dog**  
The Spanish Water Dog is a ancient and historic breed. It was first mentioned in historical literature in 1110 AD. It is also known as the Spanish Water Dog.

And the mix testo-immagini?

◆ \*US



http://www.corp.google.com/~anneaula/studies/fu-u-oct2008/searchstudygenerator/htmls/howtobeatie.html?q=how+to+tie+a+tie

File Edit View Favorites Tools Help

Page Tools

Web Images Maps News Shopping Gmail more

Sign in

# Google

how to tie a tie

Search Advanced Search Preferences

Web Images Video Results 1 - 10 of about 51,200,000 for how to tie a tie. (0.12 seconds)

**Tie-a-Tie.net | Learn How to Tie a Tie**  
Learn how to tie a tie with the Windsor, Half Windsor, Four in Hand and Pratt tie knots by following colored diagrams from Tie-a-Tie.net!  
[www.tie-a-tie.net/](http://www.tie-a-tie.net/) - 10k · Cached · Similar pages

Windsor Knot      Knot Video  
Half Windsor Knot      Tips & Tricks  
Four in Hand      Job Interview Dress  
Pratt Knot      Necktie Accessories

More results from tie-a-tie.net ...

**Tie-a-Tie.net | Windsor Knot**  
Learn how to tie a tie with the Windsor Knot, following clear instructions and colored diagrams!  
[www.tie-a-tie.net/windsor.html](http://www.tie-a-tie.net/windsor.html) - 12k · Cached · Similar pages

**How To Tie A Tie - Presented by ScoutDB.org**  
Covers the four basic knots. Diagrams and instructions.  
[www.scoutdb.org/h2tat/](http://www.scoutdb.org/h2tat/) - 5k · Cached · Similar pages

**Image results for how to tie a tie** - Report images



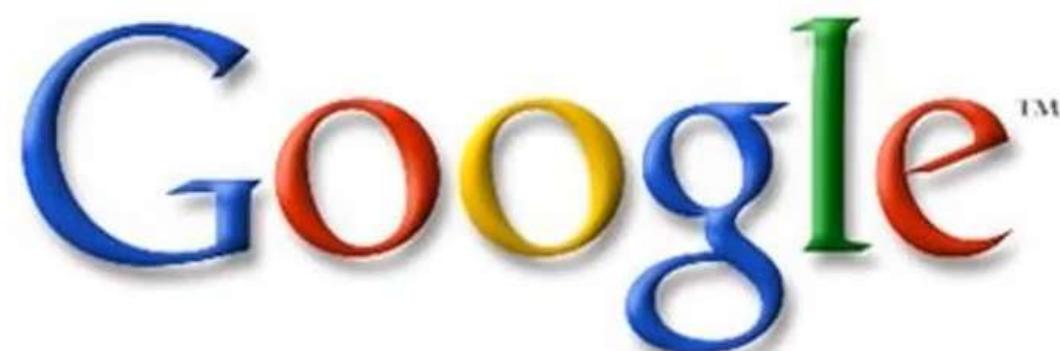
Windsor knot site: tie-a-tie.net

Internet 100% 9:30 AM

start bakery san francisco... how to tie a tie - Goog...

# Consequence

- ◆ Mixing images to text doesn't alter the properties of the «textual» top ten



http://images.google.com/images?hl=en&sa=Search+Results+for+How+to+Tie+a+Tie&um=1&ie=UTF-8&sa=X&q=Image\_result\_group%2fresnum=10

File Edit View Favorites Tools Help

Page Tools

How to tie a tie  
1599 x 1273 - 26k - gif  
[www.weblogcartoons.com](http://www.weblogcartoons.com)

... right is a display on how to tie ...  
324 x 432 - 14k - gif  
[members.aol.com](http://members.aol.com)

To tie ...  
484 x 363 - 80k - jpg  
[www.how-to-tie-a-tie-video.com](http://www.how-to-tie-a-tie-video.com)

The How To Tie A Tie Tie  
594 x 382 - 52k - jpg  
[cubeme.com](http://cubeme.com)

How to tie a tie (Windsor Knot)  
360 x 360 - 20k - jpg  
[i.ehow.com](http://i.ehow.com)

HOW TO TIE A NECKTIE WINDSOR  
STYLE.  
411 x 400 - 47k - gif  
[www.tieking.com.au](http://www.tieking.com.au)

How to tie a tie by briansuda.  
500 x 375 - 112k - jpg  
[www.flickr.com](http://www.flickr.com)

The bow tie is just so adorable even ...  
340 x 415 - 103k - jpg  
[www.urbambloc.com](http://www.urbambloc.com)

Google ►  
1 2 3 4 5 6 7 8 9 10 [Next](#)

New! Want to help improve Google Image Search? Try [Google Image Labeler](#).

how to tie a tie [Search Images](#) [Search the Web](#)

# So how do we climb the ranking?

- ◆ **SPAMDEX** = SPAM INDEX
- ◆ Also called  
**SEO** (Search Engine Optimization), or  
**SEP** (Search Engine Persuasion)

# Let's see an example...

◆ \*groovymovies

**Web  
Jump.. Free Web Hosting**



Coming Soon.....



**Groovymovies.com**

Reviews!  
Video!  
DVD!  
Current Films!  
More!

# How is rank calculated?

- ◆ Currently, information is given by the *textual* component of a page, plus its *hypertextual* component

# On hypertextual...

- ◆ You probably heard already something about it (pagerank etc): we will go back to this soon

# Let's focus before...

- ◆ On the «less famous» part, the textual one
- ◆ And let's try to see things from the search engine perspective...



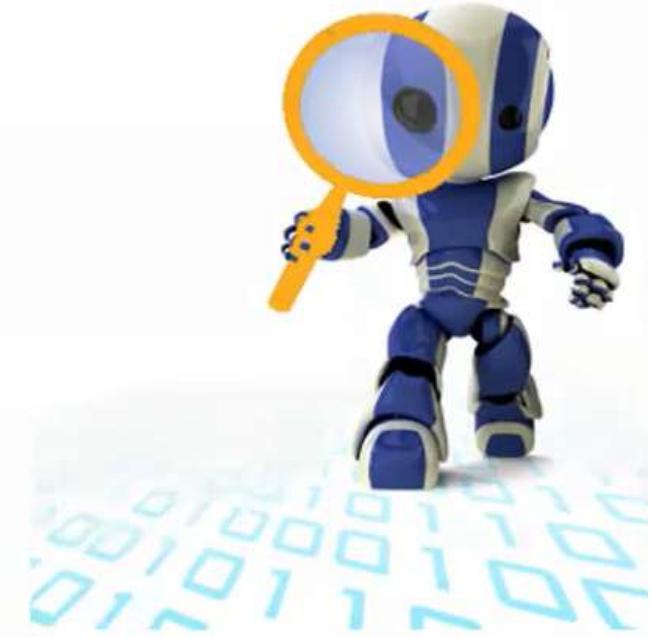
# The textual part

- ◆ All search engines use at base level variants of the same technique
- ◆ Called TFIDF (or TF-IDF)



# TFIDF

- ◆ Stands for  
**Term Frequency**  
**Inverse Document Frequency**
- ◆ Gives a measure of how important a word is for the page
- ◆ **TFIDF = TF \* IDF**



# TF



- ◆ ***Term Frequency***: the frequency of how much the word appears in the page
- ◆ Example: a web page → 100 words
- ◆ Within the page, the word «football» appears five times
- ◆ → TF of football for the page is 5% (0.05)

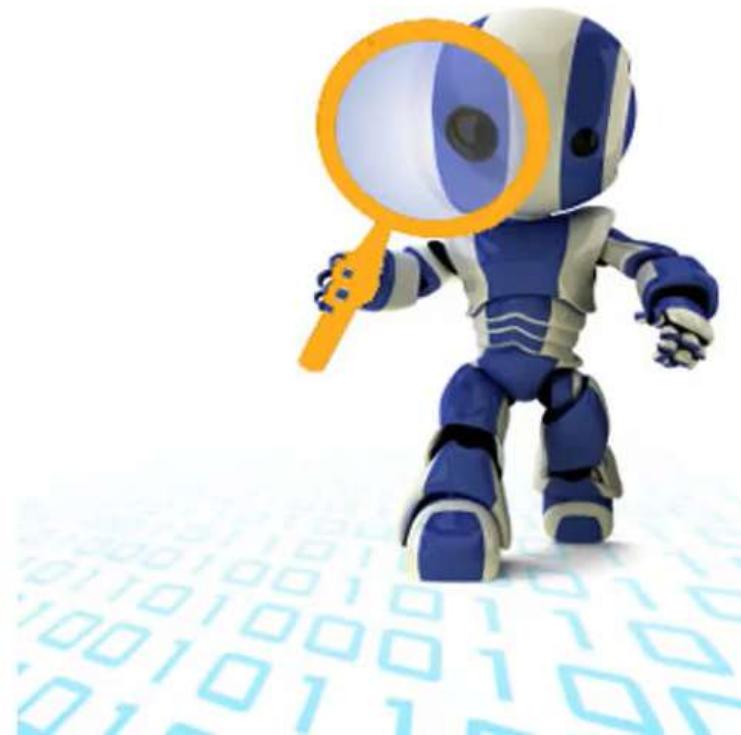
# The problem with TF

- ◆ IF we were to use TF only, many cases just would not work
- ◆ Think for instance words like «the»: they would be the most important words in a page!
- ◆ To solve this, the second component (IDF) comes to help



# IDF

- ◆ Inverse Document Frequency →  
the inverse of the frequency of the  
word **within** the set of documents  
(web site, www etc)...
- ◆ scaled logarithmically



# Example



- ◆ Web site of 1000 pages, «the» appears in 980 pages → 98% frequency (0.98) → IDF is  $\log(1/0.98) = \mathbf{0.008}$
- ◆ Web site of 1000 pages, “bike” appears in 100 pages → 10% frequency (0.1) → IDF is  $\log(1/.1) = \mathbf{1}$
- ◆ Web site of 1000 pages, “Schopenhauer” appears in 10 pages → 1% frequency (0.01) → IDF is  $\log(1/0.01) = \mathbf{10}$

# So (simplifying)

- ◆ When a search engine receives a **query** made by a word **w**, it takes all the pages where **w** appears, and then computes their TFIDF
- ◆ If the query has more than one word (word1 word2 word3 ...) we can just sum the TFIDF's computed for each word





# Note...



- ◆ The “dilemma” (tension)....:
- ◆ If we want to raise the textual score of a page for a word  $w$ , we have to be careful because raising too much its TFIDF automatically causes lowering the TFIDF of the other words!



# The strategy

- ◆ Focus on a set of words (the «champions»), and raise their TFIDF (and textual score), lowering the others



# So...

- ◆ ... we will have to carefully ***choose a set of keywords***, and appropriately ***manage*** them in the web site
- ◆ There are various ways to do it...

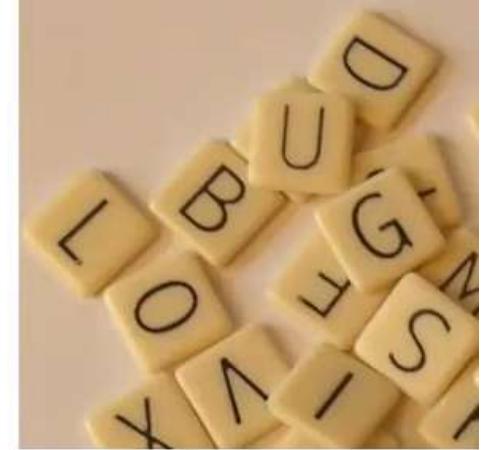


# Body spam



- ◆ The easiest way: insert words into the BODY of an HTML page
- ◆ Simple and effective (apart from the compromises with the TFIDF...)
- ◆ Disadvantage: we are touching the actual content of the page

# Title spam



- ◆ Similarly, insert the keyword in the TITLE of the page
- ◆ Advantage over body spam: the content of the page is touched way less

# Meta tag spam



- ◆ <meta name="keywords" content="bike, football, sport">
- ◆ Advantages: no (user side) visible content is touched
- ◆ Disadvantages: abused, very low score by current search engines

# Anchor text spam



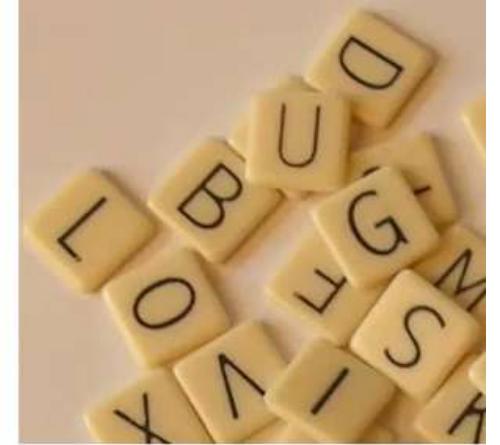
- ◆ Technically, a part of body spam, but it is usually considered apart
- ◆ Insert the words in the anchor text (<A>...</A>)
- ◆ Special scores, and a peculiarity...

# Anchor text spam (cont.)



- ◆ Breaking the (usually purely textual) model, keywords are typically also added by the search engine to the ***target*** page of the link
- ◆ And as added bonus, so with *less limitations* with respect to TFIDF!

# URL spam



- ◆ Finishing term spamming, URL spam is the technique to insert keywords directly into the web address of the page (!)
- ◆ This because search engines also analyze the addresses, giving ***bonuses*** similar to the anchor text spam

# The other side of the coin

- ◆ So far, we saw positioning in term spamming, so to say the «***form***»
- ◆ But obviously also the «***content***» matters: what keywords do we use

# The “Starter Kit”

- ◆ Repetition
- ◆ Dumping
- ◆ Weaving
- ◆ Stitching
- ◆ Broadening



# Repetition



- ◆ Repeat the same keyword,  
of course paying attention...
- ◆ ... to the TFIDF (balancing), and also to
- ◆ the **countermeasures** (!),  
given that repetition spam is easy to  
identify (and penalize...!)

# Dumping



- ◆ Insert many terms that are seldom used, even if not related to the page (!)
- ◆ → being rare keywords their score will be relative high!

# Weaving



- ◆ Take pieces of other web sites, and modify them by inserting our keywords (usually in a random way)
- ◆ So, automatic way to create interesting content, and power it up in attraction with our specific keywords



# Stitching

- ◆ Paste & copy of fragments of other web pages, uniting them into a single page
- ◆ → automatic way to create «interesting» content to populate a site (various search engine also use ***global bonuses*** to measure how much ***information*** does a site offer)