BOOLEAN INFORMATION RETRIEVAL

The *Boolean model* of *IR* (*BIR*) is

- a classical *IR* model and, at the same time,
- the first and most adopted one.

It is used by virtually all commercial *IR* systems today.

The *BIR* is based on:

- Boolean Logic and
- classical Sets Theory

in that both the documents to be searched and the user's query are conceived as sets of terms.

Retrieval is based on whether or not the documents contain the query terms.

BOOLEAN LOGIC

Proposition

The *proposition* is a statement (formulation, assertion) which can be assigned either a value T or a value F (there is no third alternative), where T and F are two different values, i.e., $T \neq F$. For example, T = true, F = false (these values are used throughout this book), or T = yes, F = no, or T = white, F = black, or $T = 1$, $F = 0$. The values T and F will be referred to as *truth values*. A proposition cannot be true and false at the same time (*principle of noncontradiction*).

Example

- "*I am reading this text.*" is a — true — proposition.
- The sentence "*The sun is shining*." is also a proposition because either the value T or F can be assigned to it.
- The sentence „*The cooks wearing red hats are playing football at the North Pole*.“ becomes a proposition if a truth value can be assigned to it.

In general, it is not a necessary quality of a proposition to be true. For example, the proposition "*It is raining*." may be true or false. However, there are propositions that are absolutely true (e.g., "The year 2001 is the first year of the 21st century.")

Negation

The *negation* of a proposition P is a proposition denoted by $\neg P$ and pronounced "not P". If P is true, then $\neg P$ is false, and if P is false, then $\neg P$ is true.

Hence, $\neg(\neg P)$ is always true (*law of double negation*).

Truth table of the logical negation

P	$\neg P$
T	F
F	T

Example

"*I am not reading this text*." is a — false — proposition, and is the negation of the proposition "*I am reading this text*.". □

Conjunction

Given two propositions: P, Q. The proposition $P \wedge Q$ (pronunciation: "P and Q") is called *conjunction*. The conjunction is true if and only if both P and Q are true, and false otherwise (Table 2.2).

Thus, $P \wedge (\neg P)$ is always false (*law of contradiction*).

Table 2.2. Truth table of the logical conjunction

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

Example 2.3.

- "*I am reading this text*. ∧ *It is raining*." is a proposition, and its truth value can be assigned by the Reader.
- "*I am thinking at myself*. ∧ *A bicycle has two wheels*." is a proposition (the Reader can assign a truth value to it), albeit one would rarely link its two constituent propositions into one sentence in everyday speech. □

Disjunction

Given two propositions: P, Q. The proposition $P \vee Q$ (pronunciation: "P or Q") is called *disjunction*. The disjunction is false if and only if both P and Q are false, and true otherwise (Table 2.3). Thus, $P \vee (\neg P)$ is always true (*law of excluded third*).

Table 2.3. Truth table of the logical disjunction

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

Example 2.4.

"*I am reading this text*. ∨ *It is raining*." is a true proposition (regardless of whether it is actually raining or not). □

SETS THEORY

Set

The notion of set is a fundamental one. It does not have a mathematical definition. A *set* is a collection of distinct objects. The objects in a set are called *elements*. If an object x is an *element of* a set S (an equivalent formulation is: *x belongs to S*), this is denoted as $x \in S$. $x \notin S$ means that x *does not belong* to S.

It is very important to note that:

- An element can occur at most once in a set.
- The order of the elements in a set is unimportant.

A set can be given by enumerating its elements between brackets, e.g., $A = \{a_1, a_2,...,a_n\}$, or by giving a property $P(x)$ (e.g., using a predicate $P(x)$; see section 2.3.3 for the notion of predicate) all elements must share as follows: $A = \{x \mid P(x)\}$. A set having a fixed number of elements is *finite*, and *infinite* otherwise. The *empty set* contains no elements and is denoted by $\varnothing$.

Example 2.7.

- $\mathbb{N} = \{1, 2,\ldots,n,\ldots\}$ denotes the set of natural numbers.
- $\mathbb{Z} =\{..., -2, -1, 0, 1, 2, ...\}$ denotes the set of integer numbers.
- $\mathbb{Q}$ denotes the set of rational numbers,
- $\mathbb{R}$ denotes the set of real numbers,
- $\mathbb{C}$ denotes the set of complex numbers,
- {thought, ape, quantum, Rembrandt} is a set.
- {mammal | water content is less than 20%} is a set. □

Union

The *union* of sets A and B is denoted by the symbol $\cup$ and defined as follows (Fig. 2.2):

$$A \cup B = \{x \mid (x \in A) \lor (x \in B)\}.$$

Example 2.10.

{thought, ape, quantum, Rembrandt} ∪ {1, 2} = {thought, ape, quantum, Rembrandt, 1, 2}. Note that the operation of union is a purely formal one (just like the other set operations): it does not require that the elements of the sets be compatible with each other, or have the same nature, in any way. □

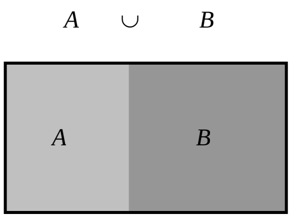


Fig. 2.2. Visualisation of the notion of set union

Set union is (as can be easily checked using the definitions of sets equality and union)

- Commutative: $A \cup B = B \cup A$, for any two sets A, B;
- Associative: $A \cup (B \cup C) = (A \cup B) \cup C$, for any three sets A, B, C;
- Idempotent: $A \cup A = A$, for any set A.

Intersection

The *intersection* of sets A and B is denoted by the symbol ∩ and defined as follows (Fig. 2.3):

$$A \cap B = \{x \mid (x \in A) \wedge (x \in B)\}.$$

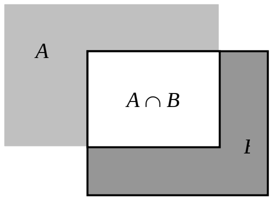


Fig. 2.3. Visualisation of the notion of set intersection

If $A \cap B = \varnothing$, the sets are said to be *disjoint* sets (Fig. 2.4).

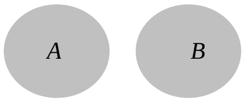


Fig. 2.4. Visualisation of the disjoint sets A and B

Example 2.11.

{thought, ape, quantum, Rembrandt} $\cap$ {thought, Rembrandt, 1, 2} = {thought, Rembrandt}. Note that the result of the intersection consists of the elements which are exactly the same. □

Set intersection is (as can be easily checked using the definitions of sets equality and intersection)

- Commutative: $A \cap B = B \cap A$, for any two sets A, B;
- Associative: $A \cap (B \cap C) = (A \cap B) \cap C$, for any three sets A, B, C;
- Idempotent: $A \cap A = A$, for any set A.

Set Difference

The *difference* of sets A and B (in this order) is denoted by the symbol \ and defined as follows (Fig. 2.5):

$$A \setminus B = \{x \mid (x \in A) \land (x \notin B)\}.$$

Example 2.12.

{thought, ape, quantum, Rembrandt} \ {thought, Rembrandt, 1, 2} = {ape, quantum}. □

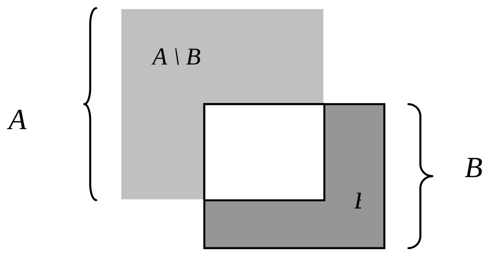


Fig. 2.5. Visualisation of the notion of set difference

Powerset

The *powerset* $\wp(A)$ of a set A is defined as follows: $\wp(A) = \{X \mid X \subseteq A\}$, i.e., the set of all subsets of A. The empty set $\varnothing$ is a member of the powerset of any set A, i.e., $\varnothing \in \wp(A)$.

Example 2.15.

$\wp$({thought, ape, quantum}) = {$\varnothing$, {thought}, {ape}, {quantum}, {thought, ape}, {thought, quantum}, {ape, quantum}, {thought, ape, quantum}}. □

Given a finite set

$$T = \{t_1, t_2, ..., t_i, ..., t_n\}$$

of elements called *terms* (e.g. words or expressions – which may be stemmed – describing or characterising documents such as keywords given for a journal article), a finite set

$$D = \{D_1, ..., D_j, ..., D_m\}, D_j \in \wp(T)$$

of elements called *documents*. Traditionally, a – real – document can be a journal article (or its abstract or title), or a newspaper article, etc. These documents for retrieval purposes, in the *BIR*, are represented as **sets of terms**.

Given a Boolean expression *Q* called a *query*, for example (the terms are *A*,*B*,*C*,*D*):

$$\neg B \wedge \neg C$$
$$(A \vee C) \wedge (B \vee C)$$
$$A \wedge (B \vee D) \wedge (B \vee E).$$

Retrieval is defined as follows:

1. The set S_i of documents are obtained that contain or not the term under focus:

   for term A: $S_i = \{D \mid A \in D\}$

   for negated term A, i.e., $\neg A$: $S_i = \{D \mid A \notin D\}$

2. Those documents are retrieved in response to Q, which belong to the set obtained as a result of the corresponding sets operations:

   intersection $\cap$ corresponds to logical AND,

   union $\cup$ corresponds to logical OR.

   For example,

$$Q = \boldsymbol{A} \textbf{ OR } (\boldsymbol{B} \textbf{ AND } \boldsymbol{C})$$

   S_1 results set for term A,

   S_2 results set for term B,

   S_3 results set for term C,

   The retrieved set in response to Q:

$$\boldsymbol{S_1 \cup (S_2 \cap S_3)}$$

EXAMPLE

Let the set of original (real) documents be, for example $O = \{O_1, O_2, O_3\}$

where

O_1 = Bayes' Principle: The principle that, in estimating a parameter, one should initially assume that each possible value has equal probability (a uniform prior distribution).

O_2 = Bayesian Decision Theory: A mathematical theory of decision-making which presumes utility and probability functions, and according to which the act to be chosen is the Bayes act, i.e. the one with highest Subjective Expected Utility. If one had unlimited time and calculating power with which to make every decision, this procedure would be the best way to make any decision.

O_3 = Bayesian Epistemology: A philosophical theory which holds that the epistemic status of a proposition (i.e. how well proven or well established it is) is best measured by a probability and that the proper way to revise this probability is given by Bayesian conditionalisation or similar procedures. A Bayesian epistemologist would use probability to define, and explore the relationship between, concepts such as epistemic status, support or explanatory power.

Let the set T of terms be:

T = {

t_1 = Bayes' Principle,

t_2 = probability,

t_3 = decision-making,

t_4 = Bayesian Epistemology

}

The set D of documents is as follows:

$D = \{D_1, D_2, D_3\}$, where

D_1 = {Bayes' Principle, probability}

D_2 = {probability, decision-making}

D_3 = {probability, Bayesian Epistemology}

Let the query Q be:

Q = probability AND decision-making

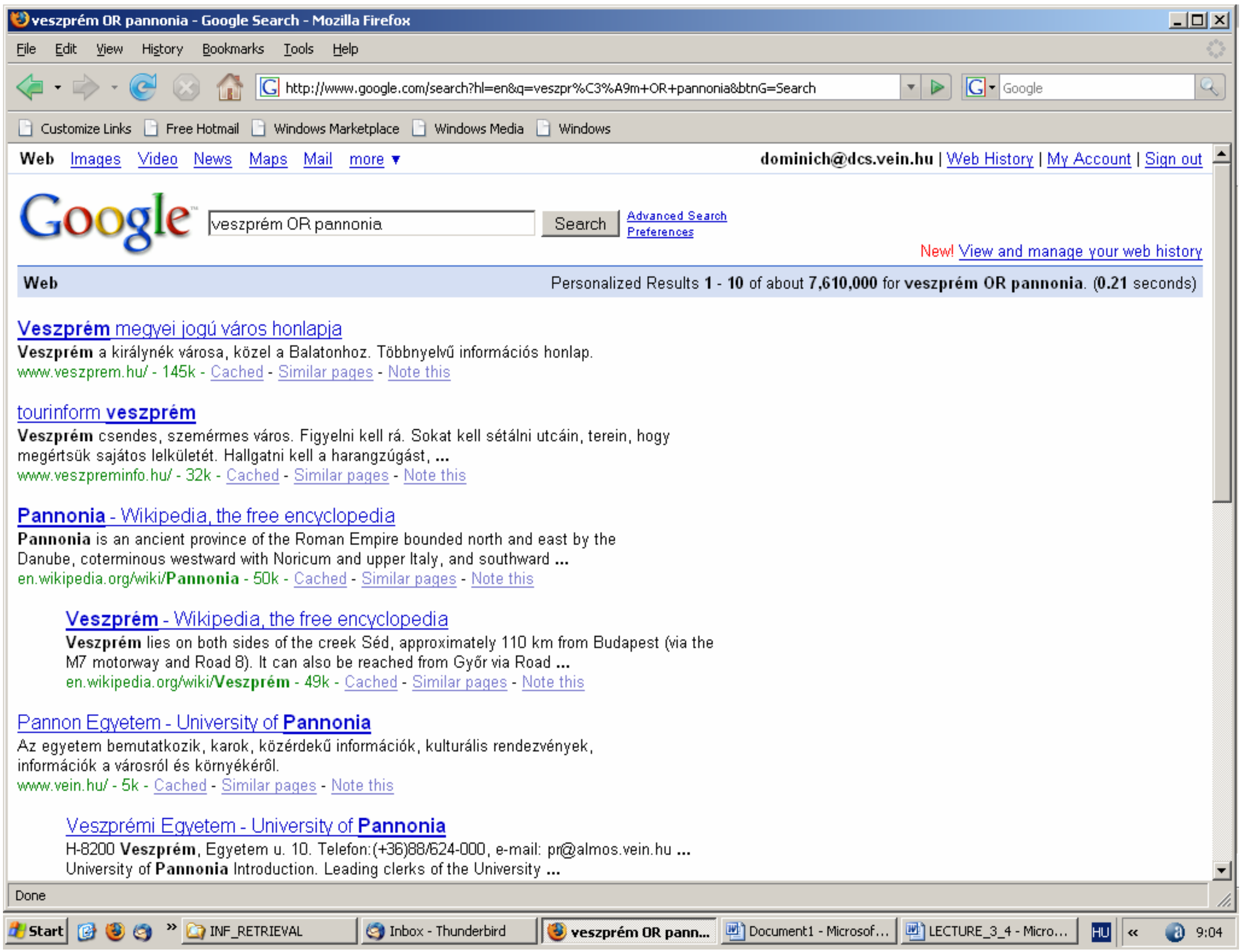
1. Firstly, the following sets S_1 and S_2 of documents D_j are obtained (retrieved):

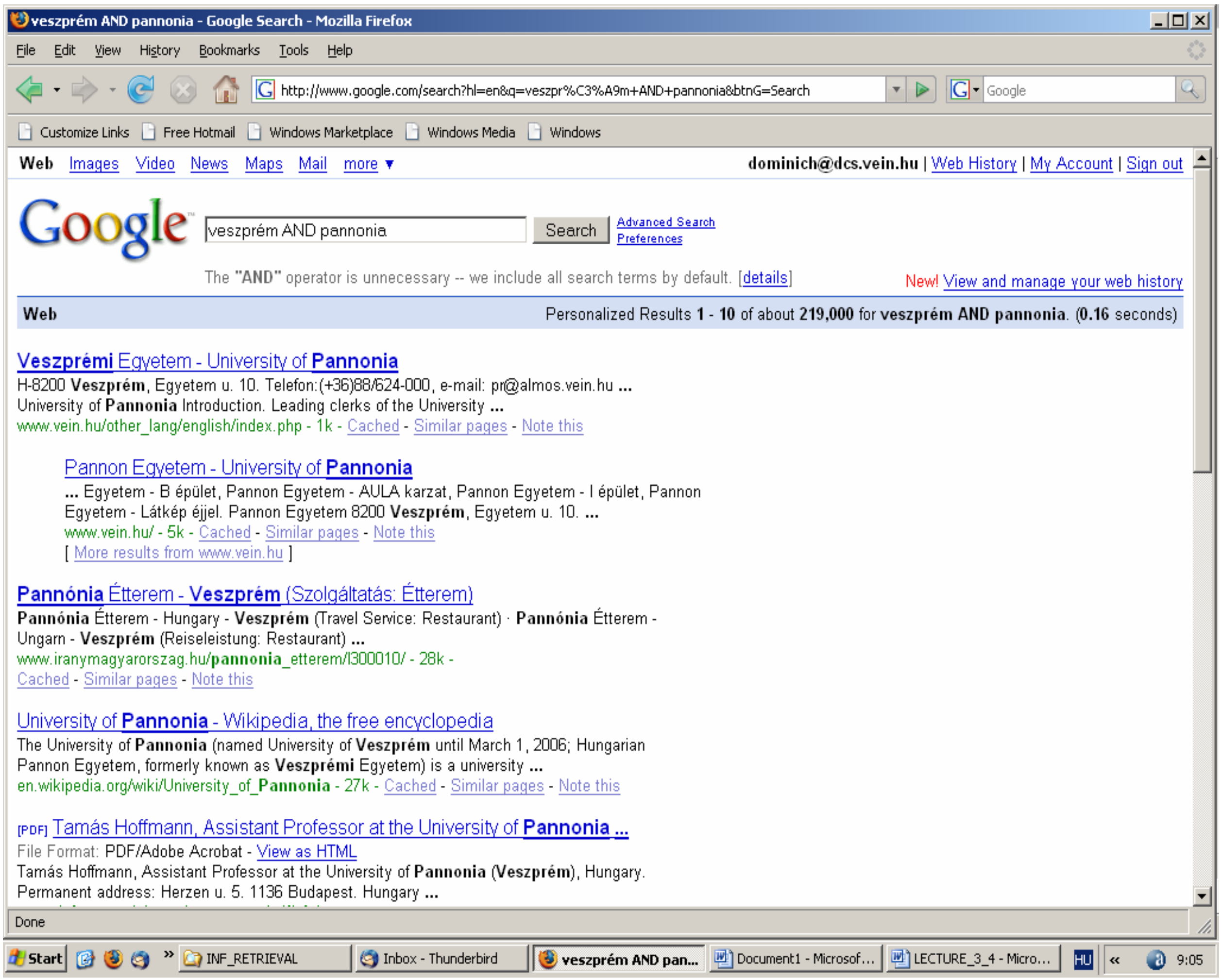
$S_1 = \{D_j |\, \text{probability} \in D_j\} = \{D_1, D_2, D_3\}$

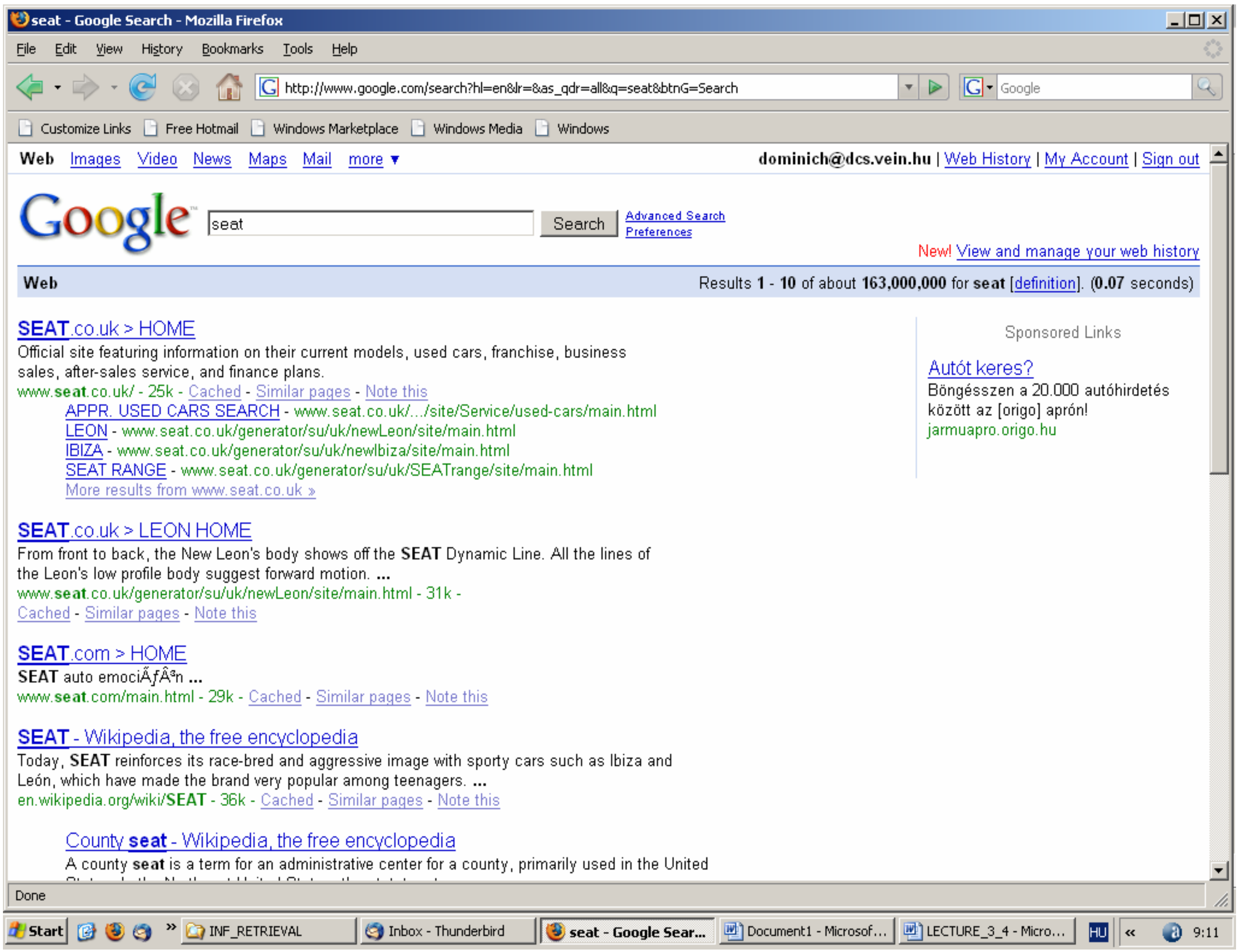
$S_2 = \{D_j |\, \text{decision-making} \in D_j\} = \{D_2\}$

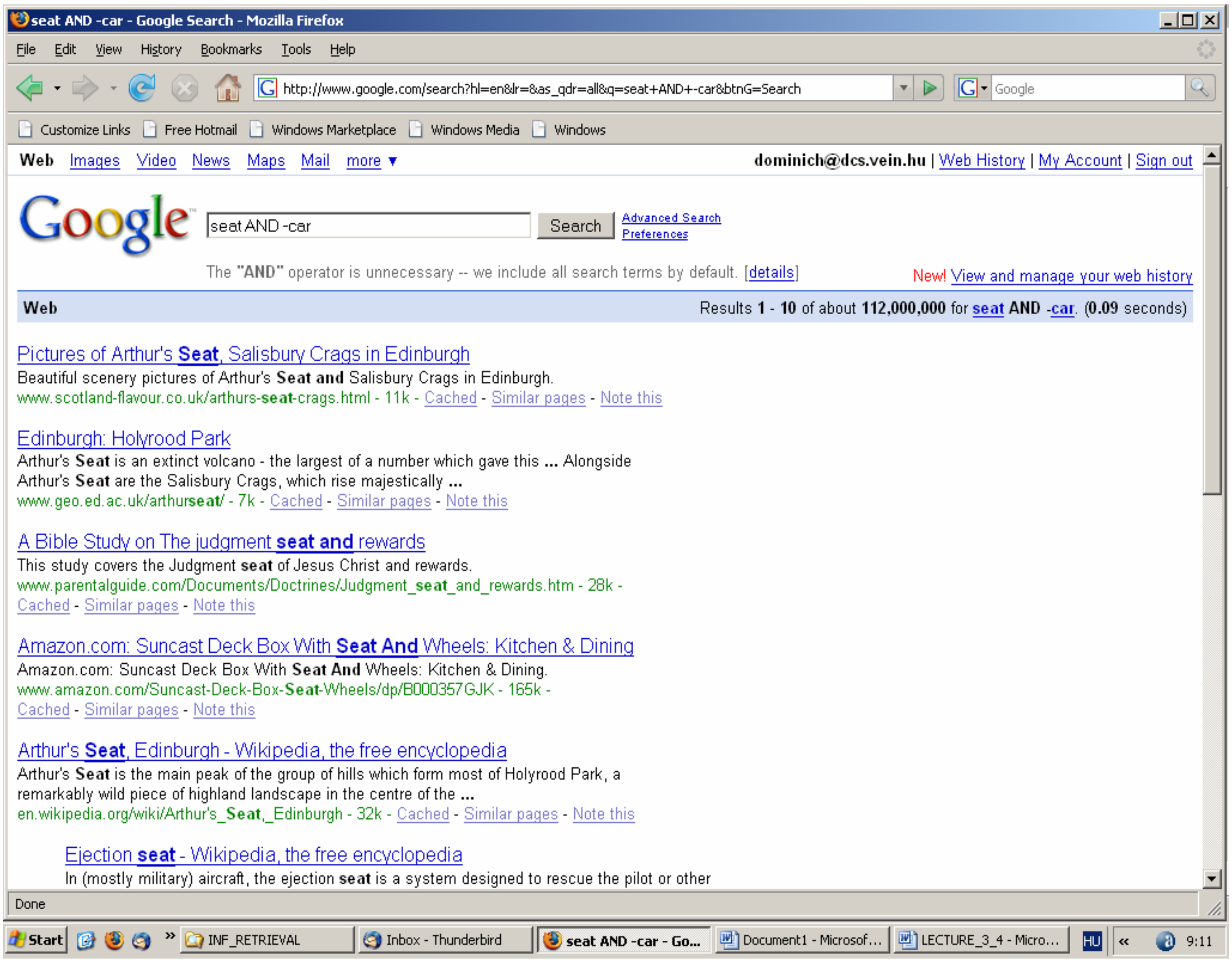
2. Finally, the following documents D_j are retrieved in response to Q:

$\{D_j \mid D_j \in S_1 \cap S_2\} = \{D_1, D_2, D_3\} \cap \{D_2\} = \{D_2\}$


veszprém OR pannonia - Google Search - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com/search?hl=en&q=veszpr%C3%A9m+OR+pannonia&btnG=Search
Customize Links Free Hotmail Windows Marketplace Windows Media Windows
Web Images Video News Maps Mail more
dominich@dcs.vein.hu | Web History | My Account | Sign out
veszprém OR pannonia
Search
Advanced Search
Preferences
New! View and manage your web history
Web
Personalized Results 1 - 10 of about 7,610,000 for veszprém OR pannonia. (0.21 seconds)
Veszprém megyei jogú város honlapja
Veszprém a királynék városa, közel a Balatonhoz. Többnyelvű információs honlap.
www.veszprem.hu/ - 145k - Cached - Similar pages - Note this
tourinform veszprém
Veszprém csendes, szemérmes város. Figyelni kell rá. Sokat kell sétálni utcáin, terein, hogy megértsük sajátos lelkületét. Hallgatni kell a harangzúgást, ...
www.veszpreminfo.hu/ - 32k - Cached - Similar pages - Note this
Pannonia - Wikipedia, the free encyclopedia
Pannonia is an ancient province of the Roman Empire bounded north and east by the Danube, coterminous westward with Noricum and upper Italy, and southward ...
en.wikipedia.org/wiki/Pannonia - 50k - Cached - Similar pages - Note this
Veszprém - Wikipedia, the free encyclopedia
Veszprém lies on both sides of the creek Séd, approximately 110 km from Budapest (via the M7 motorway and Road 8). It can also be reached from Győr via Road ...
en.wikipedia.org/wiki/Veszprém - 49k - Cached - Similar pages - Note this
Pannon Egyetem - University of Pannonia
Az egyetem bemutatkozik, karok, közérdekű információk, kulturális rendezvények, információk a városról és környékéről.
www.vein.hu/ - 5k - Cached - Similar pages - Note this
Veszprémi Egyetem - University of Pannonia
H-8200 Veszprém, Egyetem u. 10. Telefon:(+36)88/624-000, e-mail: pr@almos.vein.hu ...
University of Pannonia Introduction. Leading clerks of the University ...
Done
Start INF_RETRIEVAL Inbox - Thunderbird veszprém OR pann... Document1 - Microsof... LECTURE_3_4 - Micro... HU 9:04


veszprém AND pannonia - Google Search - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com/search?hl=en&q=veszpr%C3%A9m+AND+pannonia&btnG=Search
Customize Links Free Hotmail Windows Marketplace Windows Media Windows
Web Images Video News Maps Mail more
dominich@dcs.vein.hu | Web History | My Account | Sign out
veszprém AND pannonia
Search
Advanced Search
Preferences
The "AND" operator is unnecessary -- we include all search terms by default. [details]
New! View and manage your web history
Web
Personalized Results 1 - 10 of about 219,000 for veszprém AND pannonia. (0.16 seconds)
Veszprémi Egyetem - University of Pannonia
H-8200 Veszprém, Egyetem u. 10. Telefon:(+36)88/624-000, e-mail: pr@almos.vein.hu ...
University of Pannonia Introduction. Leading clerks of the University ...
www.vein.hu/other_lang/english/index.php - 1k - Cached - Similar pages - Note this
Pannon Egyetem - University of Pannonia
... Egyetem - B épület, Pannon Egyetem - AULA karzat, Pannon Egyetem - I épület, Pannon Egyetem - Látkép éjjel. Pannon Egyetem 8200 Veszprém, Egyetem u. 10. ...
www.vein.hu/ - 5k - Cached - Similar pages - Note this
[More results from www.vein.hu]
Pannónia Étterem - Veszprém (Szolgáltatás: Étterem)
Pannónia Étterem - Hungary - Veszprém (Travel Service: Restaurant) · Pannónia Étterem - Ungarn - Veszprém (Reiseleistung: Restaurant) ...
www.iranymagyarorszag.hu/pannonia_etterem/I300010/ - 28k -
Cached - Similar pages - Note this
University of Pannonia - Wikipedia, the free encyclopedia
The University of Pannonia (named University of Veszprém until March 1, 2006; Hungarian Pannon Egyetem, formerly known as Veszprémi Egyetem) is a university ...
en.wikipedia.org/wiki/University_of_Pannonia - 27k - Cached - Similar pages - Note this
[PDF] Tamás Hoffmann, Assistant Professor at the University of Pannonia ...
File Format: PDF/Adobe Acrobat - View as HTML
Tamás Hoffmann, Assistant Professor at the University of Pannonia (Veszprém), Hungary.
Permanent address: Herzen u. 5. 1136 Budapest. Hungary ...
Done
Start INF_RETRIEVAL Inbox - Thunderbird veszprém AND pan... Document1 - Microsof... LECTURE_3_4 - Micro... HU 9:05


seat - Google Search - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com/search?hl=en&lr=&as_qdr=all&q=seat&btnG=Search
Customize Links Free Hotmail Windows Marketplace Windows Media Windows
Web Images Video News Maps Mail more
dominich@dcs.vein.hu | Web History | My Account | Sign out
seat
Search
Advanced Search
Preferences
New! View and manage your web history
Web
Results 1 - 10 of about 163,000,000 for seat [definition]. (0.07 seconds)
SEAT.co.uk > HOME
Official site featuring information on their current models, used cars, franchise, business sales, after-sales service, and finance plans.
www.seat.co.uk/ - 25k - Cached - Similar pages - Note this
APPR. USED CARS SEARCH - www.seat.co.uk/.../site/Service/used-cars/main.html
LEON - www.seat.co.uk/generator/su/uk/newLeon/site/main.html
IBIZA - www.seat.co.uk/generator/su/uk/newIbiza/site/main.html
SEAT RANGE - www.seat.co.uk/generator/su/uk/SEATrange/site/main.html
More results from www.seat.co.uk »
SEAT.co.uk > LEON HOME
From front to back, the New Leon's body shows off the SEAT Dynamic Line. All the lines of the Leon's low profile body suggest forward motion. ...
www.seat.co.uk/generator/su/uk/newLeon/site/main.html - 31k -
Cached - Similar pages - Note this
SEAT.com > HOME
SEAT auto emociÃƒÂ³n ...
www.seat.com/main.html - 29k - Cached - Similar pages - Note this
SEAT - Wikipedia, the free encyclopedia
Today, SEAT reinforces its race-bred and aggressive image with sporty cars such as Ibiza and León, which have made the brand very popular among teenagers. ...
en.wikipedia.org/wiki/SEAT - 36k - Cached - Similar pages - Note this
County seat - Wikipedia, the free encyclopedia
A county seat is a term for an administrative center for a county, primarily used in the United
Sponsored Links
Autót keres?
Böngésszen a 20.000 autóhirdetés között az [origo] aprón!
jarmuapro.origo.hu
Done
Start INF_RETRIEVAL Inbox - Thunderbird seat - Google Sear... Document1 - Microsof... LECTURE_3_4 - Micro... HU 9:11


seat AND -car - Google Search - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com/search?hl=en&lr=&as_qdr=all&q=seat+AND+-car&btnG=Search
Customize Links Free Hotmail Windows Marketplace Windows Media Windows
Web Images Video News Maps Mail more
dominich@dcs.vein.hu | Web History | My Account | Sign out
seat AND -car
Search
Advanced Search
Preferences
The "AND" operator is unnecessary -- we include all search terms by default. [details]
New! View and manage your web history
Web
Results 1 - 10 of about 112,000,000 for seat AND -car. (0.09 seconds)
Pictures of Arthur's Seat, Salisbury Crags in Edinburgh
Beautiful scenery pictures of Arthur's Seat and Salisbury Crags in Edinburgh.
www.scotland-flavour.co.uk/arthurs-seat-crags.html - 11k - Cached - Similar pages - Note this
Edinburgh: Holyrood Park
Arthur's Seat is an extinct volcano - the largest of a number which gave this ... Alongside Arthur's Seat are the Salisbury Crags, which rise majestically ...
www.geo.ed.ac.uk/arthurseat/ - 7k - Cached - Similar pages - Note this
A Bible Study on The judgment seat and rewards
This study covers the Judgment seat of Jesus Christ and rewards.
www.parentalguide.com/Documents/Doctrines/Judgment_seat_and_rewards.htm - 28k -
Cached - Similar pages - Note this
Amazon.com: Suncast Deck Box With Seat And Wheels: Kitchen & Dining
Amazon.com: Suncast Deck Box With Seat And Wheels: Kitchen & Dining.
www.amazon.com/Suncast-Deck-Box-Seat-Wheels/dp/B000357GJK - 165k -
Cached - Similar pages - Note this
Arthur's Seat, Edinburgh - Wikipedia, the free encyclopedia
Arthur's Seat is the main peak of the group of hills which form most of Holyrood Park, a remarkably wild piece of highland landscape in the centre of the ...
en.wikipedia.org/wiki/Arthur's_Seat,_Edinburgh - 32k - Cached - Similar pages - Note this
Ejection seat - Wikipedia, the free encyclopedia
In (mostly military) aircraft, the ejection seat is a system designed to rescue the pilot or other
Done
Start INF_RETRIEVAL Inbox - Thunderbird seat AND -car - Go... Document1 - Microsof... LECTURE_3_4 - Micro... HU 9:11

Q = Strongest Earth quakes in the last ten years

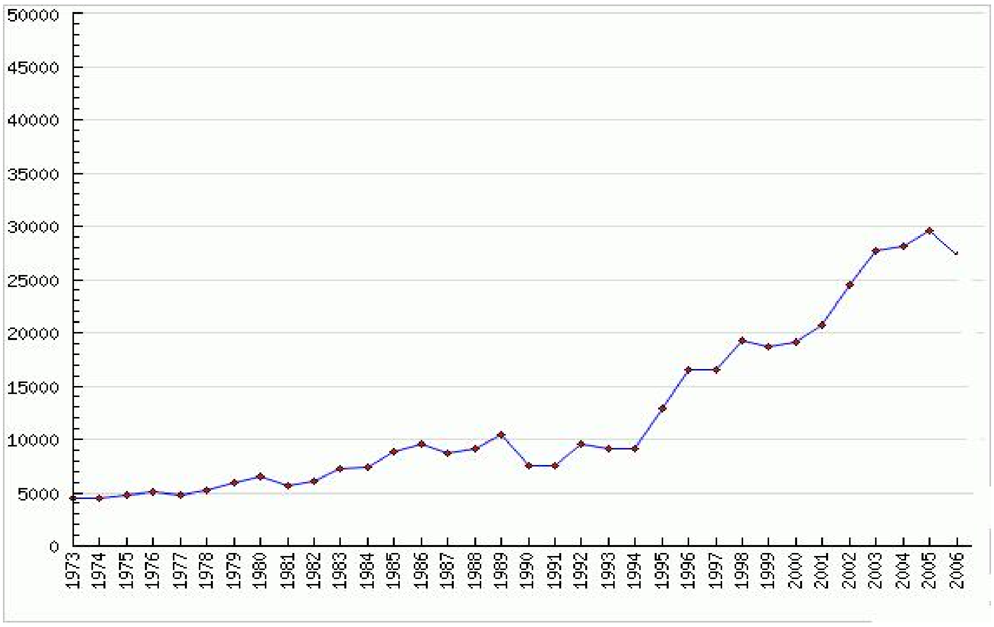
Result:
Here is a lesson using the data set of the number of earthquakes from 1900 to 1989 with a magnitude of 7 or greater. I found this after browsing the Internet at this location: United States Geological Survey (USGS), U.S. Department of the Interior. I clicked on Geologic Data, and then clicked on the National Earthquake Information Center. I found the data set after clicking on Earthquake Statistics.

Q = earth quakes magnitude 7 1997 2007

Result:

DateMagnitudeLocation February 26, 2007magnitude 1.7 approx. 7 miles north-northwest of Ellsworth February 1, 2007magnitude 2.1 approx. 5.5 miles southwest of Augusta December 29, 2006magnitude 3.1 east of Cadillac Mountain, Mount Desert Island December 18, 2006magnitude 2.3 east of Cadillac Mountain, Mount Desert Island October 22, 2006magnitude 2.3 east of
. . . . .

Q = earth AND quakes AND magnitude AND 7 AND 1997 AND 2007

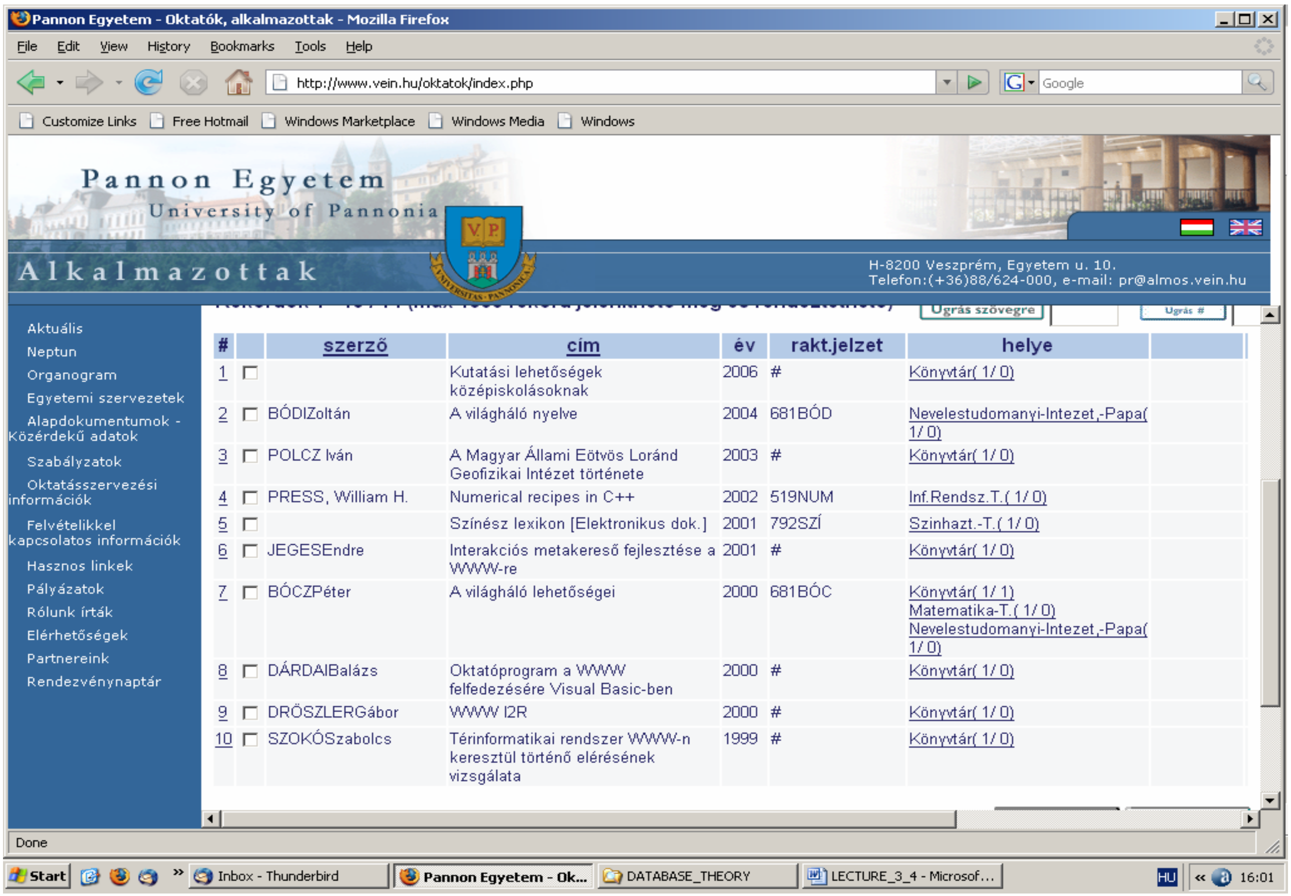


Using AND results in less results than using OR.

Q = gulyás OR paprikás; Number of hits = 1 630 000

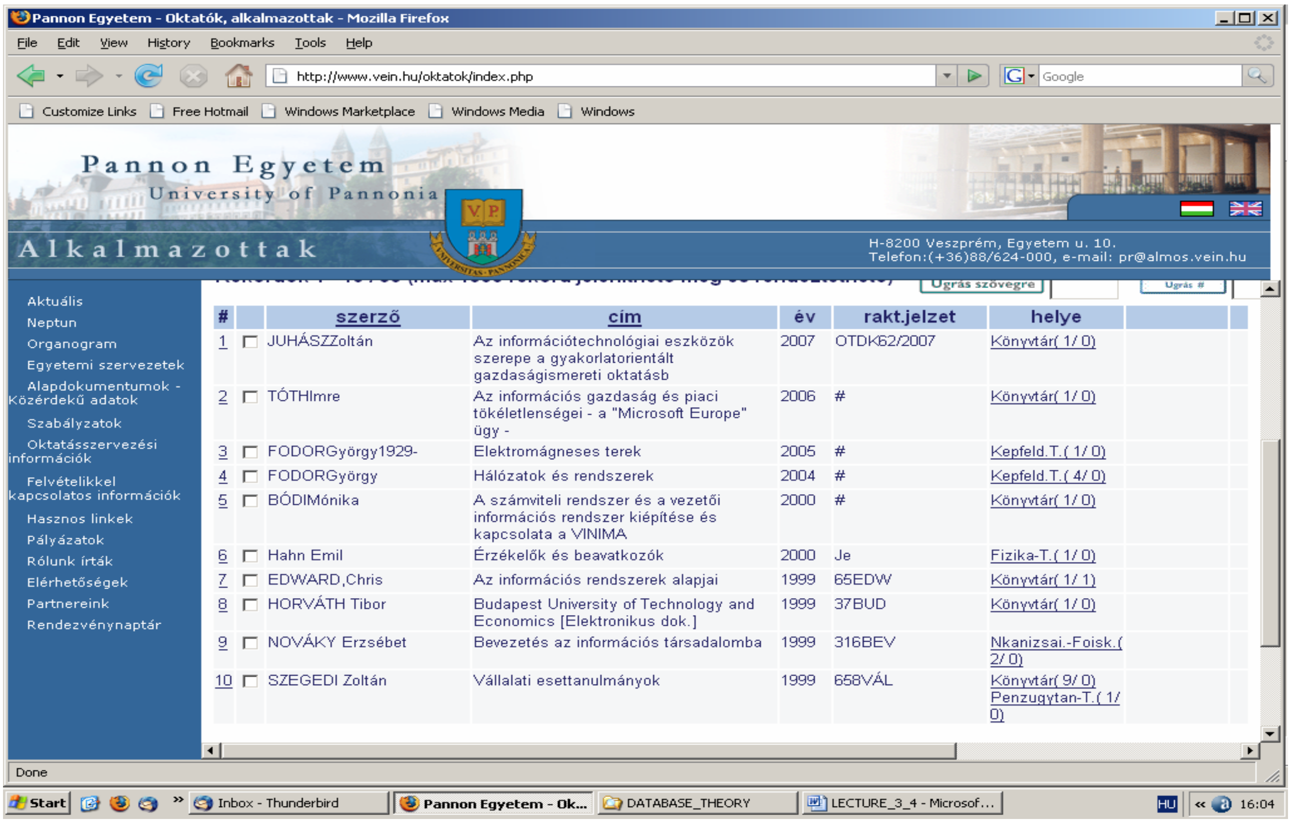
Q = gulyás AND paprikás; Number of hits = 33 600

Q = www OR világháló*



#		szerző	cím	év	rakt.jelzet	helye
1			Kutatási lehetőségek középiskolásoknak	2006	#	Könyvtár(1/ 0)
2		BÓDIZoltán	A világháló nyelve	2004	681BÓD	Nevelestudomanyi-Intezet,-Papa(1/ 0)
3		POLCZ Iván	A Magyar Állami Eötvös Loránd Geofizikai Intézet története	2003	#	Könyvtár(1/ 0)
4		PRESS, William H.	Numerical recipes in C++	2002	519NUM	Inf.Rendsz.T.(1/ 0)
5			Színész lexikon [Elektronikus dok.]	2001	792SZÍ	Szinhazt.-T.(1/ 0)
6		JEGESEndre	Interakciós metakereső fejlesztése a WWW-re	2001	#	Könyvtár(1/ 0)
7		BÓCZPéter	A világháló lehetőségei	2000	681BÓC	Könyvtár(1/ 1) Matematika-T.(1/ 0) Nevelestudomanyi-Intezet,-Papa(1/ 0)
8		DÁRDAIBalázs	Oktatóprogram a WWW felfedezésére Visual Basic-ben	2000	#	Könyvtár(1/ 0)
9		DRÖSZLERGábor	WWW I2R	2000	#	Könyvtár(1/ 0)
10		SZOKÓSzabolcs	Térinformatikai rendszer WWW-n keresztül történő elérésének vizsgálata	1999	#	Könyvtár(1/ 0)

Q = inf* AND (gazd* OR ökönöm*)



#		szerző	cím	év	rakt.jelzet	helye
1		JUHÁSZZoltán	Az információtechnológiai eszközök szerepe a gyakorlatorientált gazdaságismereti oktatásb	2007	OTDK62/2007	Könyvtár(1/ 0)
2		TÓTHImre	Az információs gazdaság és piaci tökéletlenségei - a "Microsoft Europe" ügy -	2006	#	Könyvtár(1/ 0)
3		FODORGyörgy1929-	Elektromágneses terek	2005	#	Kepfeld.T.(1/ 0)
4		FODORGyörgy	Hálózatok és rendszerek	2004	#	Kepfeld.T.(4/ 0)
5		BÓDIMónika	A számviteli rendszer és a vezetői információs rendszer kiépítése és kapcsolata a VINIMA	2000	#	Könyvtár(1/ 0)
6		Hahn Emil	Érzékelők és beavatkozók	2000	Je	Fizika-T.(1/ 0)
7		EDWARD,Chris	Az információs rendszerek alapjai	1999	65EDW	Könyvtár(1/ 1)
8		HORVÁTH Tibor	Budapest University of Technology and Economics [Elektronikus dok.]	1999	37BUD	Könyvtár(1/ 0)
9		NOVÁKY Erzsébet	Bevezetés az információs társadalomba	1999	316BEV	Nkanizsai.-Foisk.(2/ 0)
10		SZEGEDI Zoltán	Vállalati esettanulmányok	1999	658VÁL	Könyvtár(9/ 0) Penzugytan-T.(1/ 0)