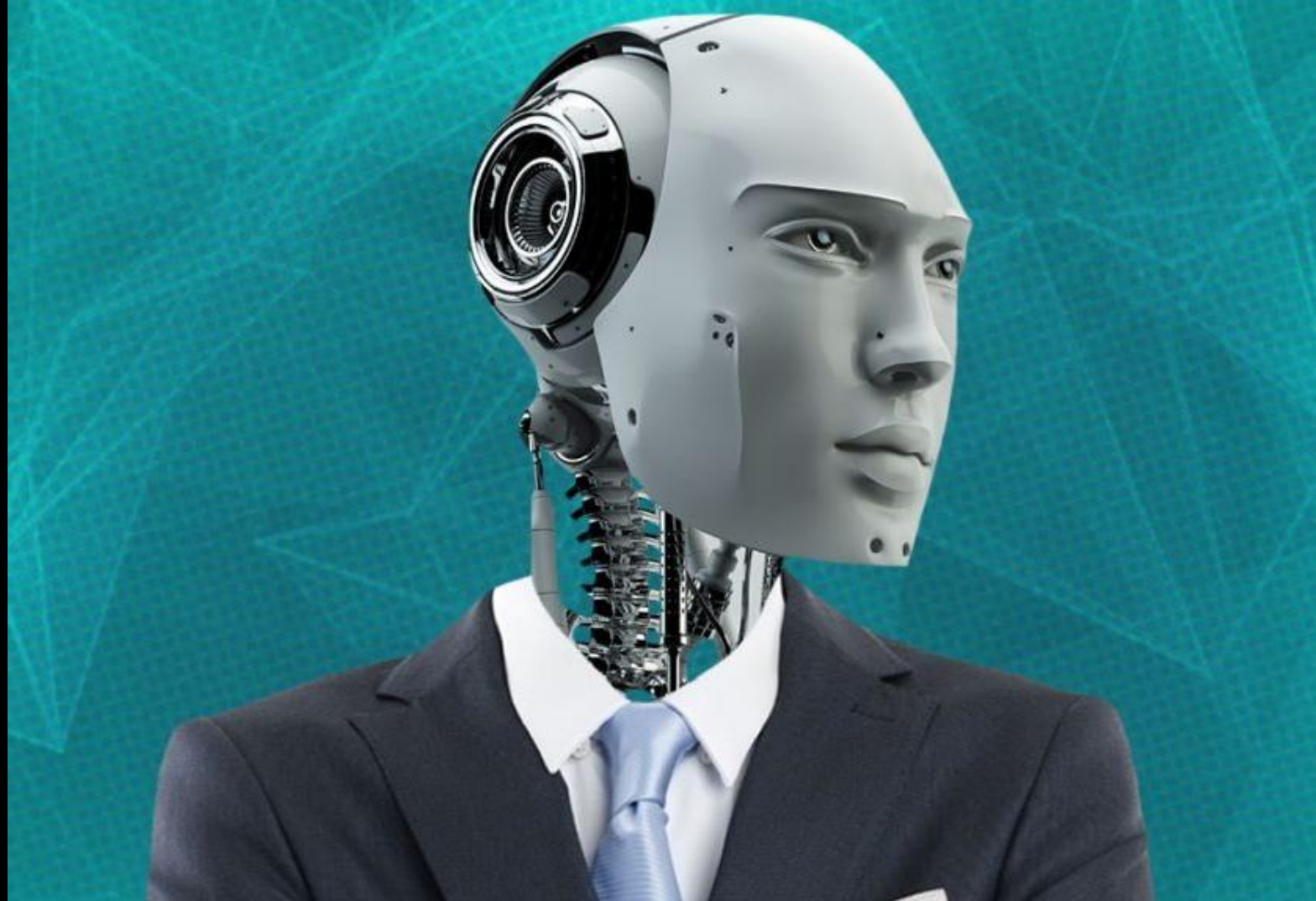


Jogi doktrína

12. Előadás

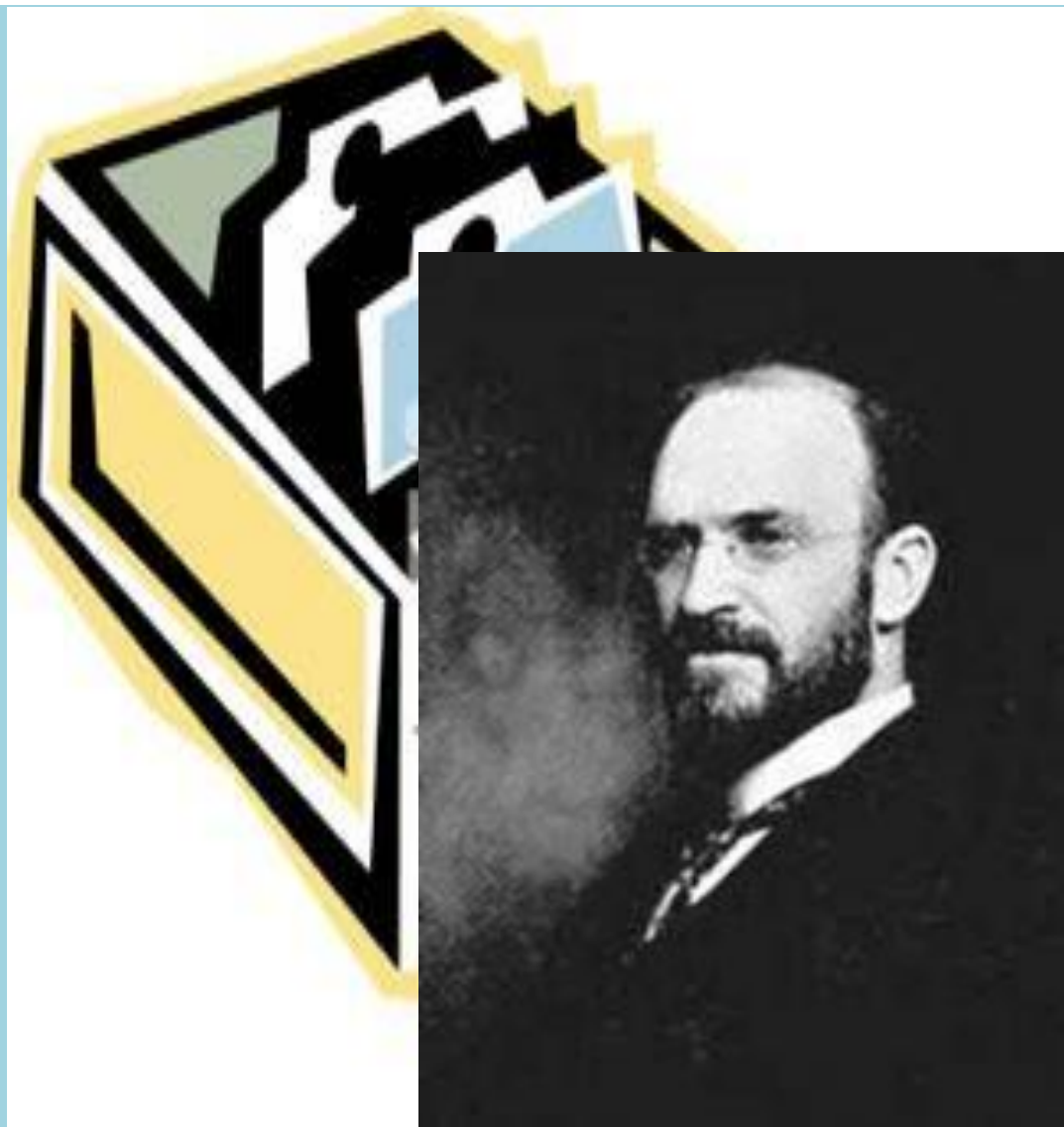
A magyarázat forrásai

Jogi adatbázisok



Tudásszervezés a jog világán kívül

- **Tudásszervezés** = módszerek, hogy egy nagy információhalmazból a nekünk szükségeset könnyebben (gyorsabban) ki tudjuk nyerni.
- **Szurrogátum** = helyettesítő, valami helyett álló. Hogy meglegjünk az információt, kisebb, helyettesítő jelekkel, jelhalmazokkal kell behelyettesítenünk a szöveg „tartalmát”, „jelentését”.
- Ilyen: a *cím*, a *tartalomjegyzék*, a *tárgymutató*, a *névmutató* (= **indexek**) – egy szövegen (könyvön) belül.
- Minél nagyobb a szöveghalmaz, annál kifinomultabb módszerek kellene.
- Pl. könyvgyűjteményekhez katalóguscédulák : a cím a szerző, kiadás éve stb. alapján történő keresés.
- Ezért születik meg a **tizedes osztályozás**, amely az összes könyvben található „téma”, azaz magának a leírt emberi tudásnak a szervezését teszi lehetővé egy egységes rendszer szerint (Melville Louis Kossuth Dewey, 1876)



Tudásszervezés: az amerikai esetjog

- **Cél:** eligazodás az ítéletek tömegében (Pl. egy hatalmas tartalomjegyzékkel).
- A jogban ez kevés, mert itt a megtalált anyag tekintélye, fontossága is számít.
- Ennek mérésére az adott ítélet későbbi említéseiről is kimutatást.
- Ez a táblázat nemcsak az említések helyét, hanem **kontextusát** is jelezte. (Pl. negatív összefüggésben, pozitív összefüggésben hivatkoztak.)

The image shows two pages from a legal research manual. The left page is a 'KEY NUMBER SYSTEM LAW CHART' with a grid of categories and sub-categories. The right page is a 'SAMPLE PAGE' from a Massachusetts case report, with various annotations and key numbers overlaid on the text.

US REQUIS	MEANING
c (criticised)	Soundness of decision or reasoning in cited case criticised for reasons given.
d (distinguished)	Case at bar different either in law or fact from case cited for reasons given.
e (explained)	Statement of import of decision in cited case. Not merely a restatement of the facts.
f (followed)	Cited as controlling.
h (harmonized)	Apparent inconsistency explained and shown not to exist.
j (dissenting opinion)	Citation in dissenting opinion.
L (limited)	Refusal to extend decision of cited case beyond precise issues involved.
o (overruled)	Ruling in cited case expressly overruled.
p (parallel)	Citing case substantially alike or on all fours with cited case in its law or facts.
q (questioned)	Soundness of decision or reasoning in cited case questioned.

Az '50-es-'60-as évek fejleménye

- University of Pittsburgh Health Law Center
- **Feladat:** az összes egészségügyi jogszabály felleltározása, és egy kifejezés általános jellegű cseréje az anyagban számítógép segítségével.
- Horthy professzor (nem 'Horthy'!) célszerű már magát az index készítését is automatizálni – ez lesz az első teljes szövegű gépi index a jog világában.

The way they worked with it was this: Looking at Fig. 2, you will see what looks like a list of words. This is just a very small portion of the vocabulary list which is composed by the machine from the total set of documents that went into it. That is, the machine has sorted out and arranged alphabetically words which appear in all of the documents with one exception -- we predetermined a list of common words -- as you can see, a list of thirty, more or less. These words are "of", "and", "the", "offer", and "begin". These were words that we felt were either used so often or so unimportant that we would not want to use them in searching a vocabulary list.

partially	339				
particular	63	151	152	219	294
parties	66	122	399		
partly	368	386	391	412	
partnership	170	270			
partnerships	66	77			
parts	128	362			

Fig. 2

After each word there are numbers. These are the numbers of the documents in which these words appear. For example, the word "particular" appears in Document 63, Document 151, Document 152, Document 219, and Document 294. This enabled us to have a list of vocabulary which not only listed the words in the total set of vocabulary but also told where each word was used so that if we were to put into the machine "particular" we could get out either the citations or full texts of all the statutes in which the word "particular" was used.



Az adatbázis

- **Adatbázis:** adatok rendezett halmaza.
- **Adat:** lehet numerikus, vagy szöveges adat.
- **Adatbázis:** több, mint egy egyszerű adathalmaz, rend van benne.
- **Rend:** az adatok csoportokba rendezhetőek, véges számú típusba sorolhatóak, és ezen belül véges számú értéket vehetnek fel.
- **Az adatbázis célja:** a tudás koncentrációja, és egy meghatározott „tétel” gyorsabb megtalálásának elősegítése.
- **Tétel = információ.** Lehet egy fizikai tétel (pl. könyv, alkatrész, áru, vagy irat), de lehet „nem megfogható” is, további adat, vagy adatok egy meghatározott, egymásra vonatkozó halmaza.

Relációs adatbázis

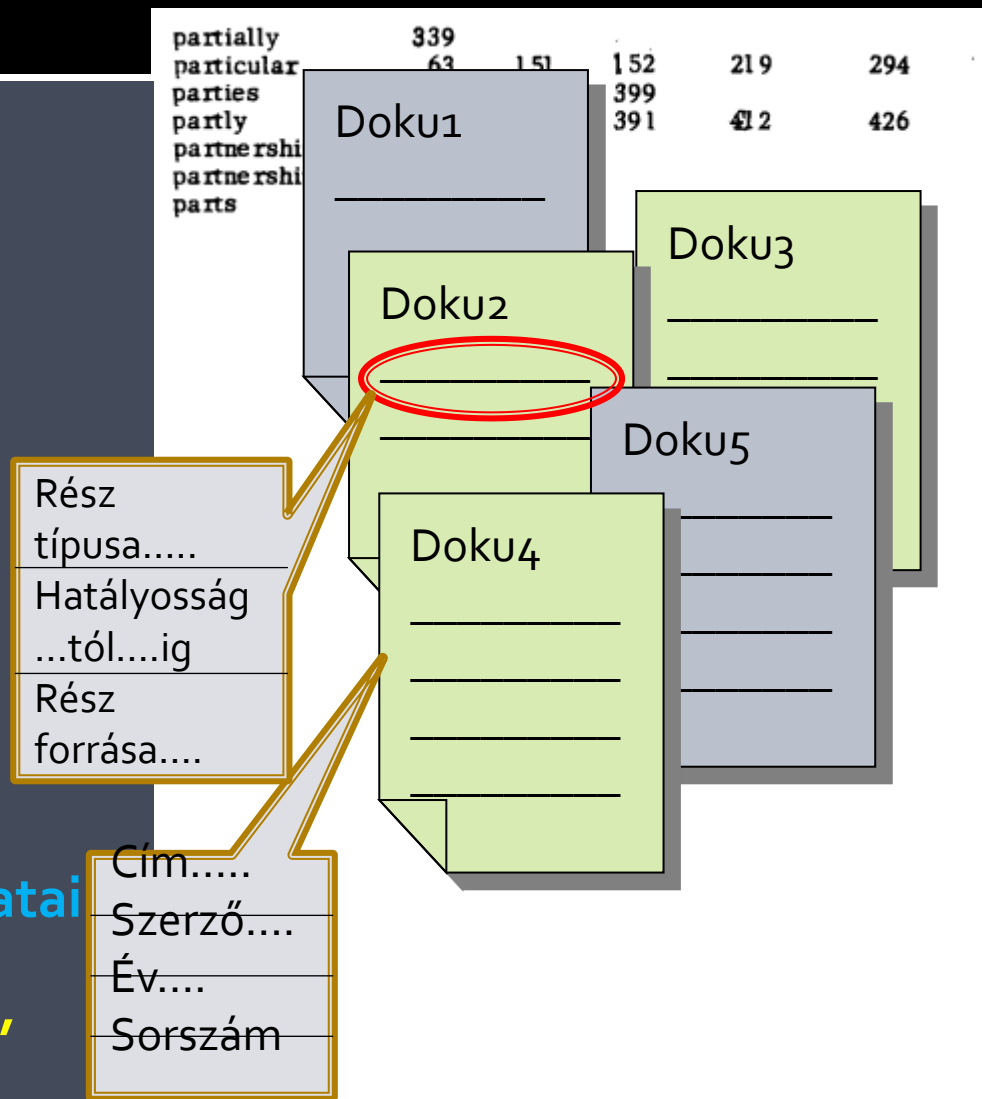
- A táblázatok az adatbázis adattáblái
- A sorokban a „rekordok”, míg az oszlopokban a „mezők” szerepelnek.
- Számuk tetszőlegesen szaporítható
- Az információ lekérdezéssel nyerhető ki belőlük, amely egy vagy több táblázat adatainak összekapcsolását jelenti.
- Pl. hány könyv jelent meg 1991-ben, vagy mikor jelentek meg Bécsben született szerzők könyvei.

1	A jog szerkezete	Pokol Béla	Gondolat	1991
2	Nikomakhoszi etika	Arisztotelész	Magyar Helikon	1971
3	Végzetes önhittség	F.A. Hayek	Tankönyvkiadó	1992
4	Út a szolgasághoz	F.A. Hayek	Közgazdasági és jogi könyvkiadó	1991

Szerző	Könyvek sorszáma amelyeket írt	Nemzetisége	Születési ideje	Születésének helye
Arisztotelész	2	görög	kr. e. 384	Stageira
F.A. Hayek	3,4	osztrák	kr. u. 1899	Bécs
Pokol Béla	1	magyar	kr. U. 1950	Záhony

Szöveges adatbázisok

- A szöveges adatbázisoknak két része van:
 - A dokumentumok
 - Az indexek
 - A teljes szövegű index
 - Az „adatlapok”
- Az adatlapokat sem mindig kézzel töltik ki – egy részüket a gép is ki tudja tölteni.
- A dokumentumokat fel is lehet darabolni > az egyes daraboknak is lehet adatlapja
- A dokumentumokat nagyobb csoportokba is lehet szervezni > ezeknek is lehetnek adatai
- Mintha címkéket ragasztanánk a lapokra, azok egyes részleteire, vagy a „dossziéokra”
- Ennek számítástechnikai módszere az XML nyelv (markup – jelölés = címke)



Az XML és jogi hasznosításai

Pl. <doku eleje>

```
<cím>  
  <évszám>  
    2012  
  <évszám vége>  
  <sorszám>  
    12  
  <sorszám vége>  
  <dokumentumtípus>  
    jegyzőkönyv (listából  
válaszható: jegyzőkönyv,  
feljegyzés,  
határozat)  
  <dokumentumtípus  
vége>  
  <cím vége>  
  <dátum>  
    2012. 01. 03.  
  <dátum vége>  
  <szöveg>  
    Első mondat. Második mondat.  
  <szöveg vége>
```

Harmadik mondat.

<doku vége>

- Az XML (Extensible Markup Language) egy kijelölő („címkeragasztgató” nyelv)
- A címkéken belül további címkék is lehetnek.
- Ez miért jó? Egy szabadszöveges dokumentumot darabokra lehet bontani, azon belül műveleteket lehet végezni a darabokkal (pl. meg lehet mondani az adott darab milyen értéket vehet fel, – dátum, szám, szöveges lista)
- A darabokra bontott, strukturált dokumentumot a számítógépek is „meg tudják érteni”.

Gépi szövegfeldolgozás

- A számítógép feltalálása után rögtön felvetődött, hogy annak számítási sebességét szövegek megtalálásához is lehetne használni.
- **Kétféle módszer:**
 - Használni lehet a *már meglévő, emberek által készített szöveg-sűrítményeket (szurrogátumokat)*; hátrány: **mindent el kell olvasni** előtte, és fennáll a veszélye, hogy a feldolgozó **nem jó hely(ek)re sorolja** be a szöveget, nem megfelelő tárgyszavakat rendel hozzá.
 - A gép *maga kivonatolja a szöveget*. Minden szót felleltároz, és ezek előfordulási helyeit is. > Majd ezután a maga által készített táblázatban keresi ki a szavakat. > ez a **TELJES SZÖVEGŰ GÉPI INDEX**



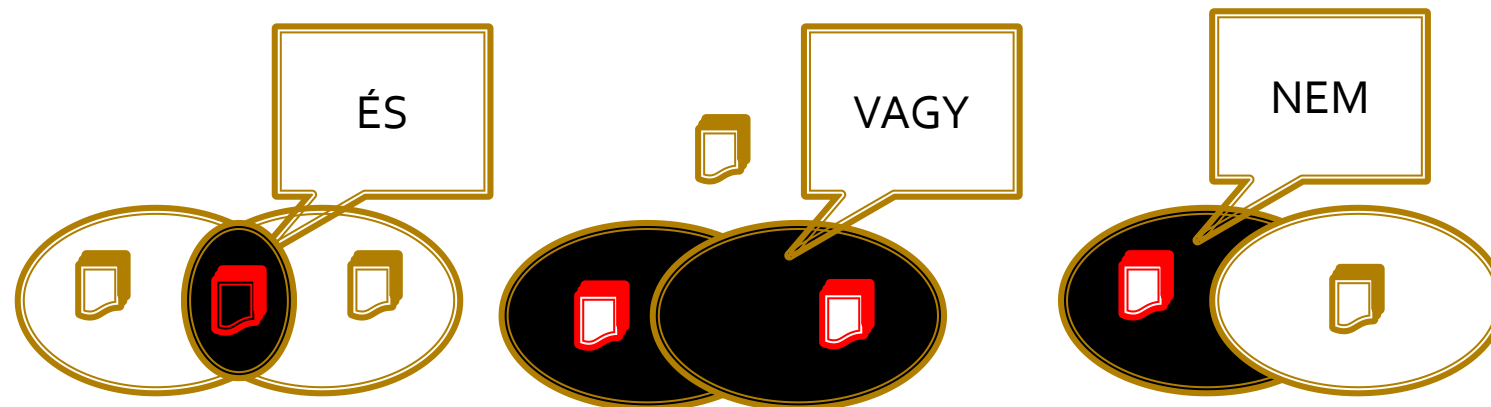
szeróvay	4
szerózus	72
szerődés	18
szerődésben	1
szerődések	1
szerődéses	2
szerődésében	1
szerődéskötés kori	3
szerődésről	1
szerődésszegése	1
szerődést	2
szerődő	5



A CD Jogtár gépi indexének egy darabja: a gép a helytelenül írt szavakat is fegyelmezetten felleltározza

A gépi index működése

- Gépi indexkészítés: a szöveg összes szavát felleltározzák, és megjelölik az előfordulási helyeiket is. (pl. szerződés – az 1, 3, 7, dokumentumokban; ha a dokumentum tovább van darabolva, az illető részt is ...)
- Amikor keresünk egy gyűjteményben, a gép kikeresi a szót a táblázatban, majd a bekezdéseket. Ezután az adott bekezdésben található összes szóból kiválasztja a keresettet, esetleg meg is jelöli – kiszínezi, 'highlight'-olja).
- Ha több szóra keresünk, mindkettő összes előfordulását kikeresi, és ezeket aszerint veti össze, ahogy ÉS, VAGY, NEM operátorokat használunk.



Hagyományos keresők

- A hagyományos szöveges adatbázisokban **kétféle módon lehet eligazodni:**
 - A könyveknél megszokott szurrogátumokon keresztül → navigáció
 - Tartalomjegyzék
 - Tárgyszólista
 - Hivatkozások
 - Gépi indexeken keresztül → keresés
 - Teljes szövegű keresés – Boole operátorok
 - Elhelyezkedési szabályok
 - Közelségi szabályok
 - Sorrend szabályok

Adatbázis; tartalom; tárgyszó

- CompLex Jogtár
- CompLex Jogtár Plusz
- CompLex határozatok, törvénykezés**
 - Európai Közösség Bíróságai
 - Alkotmánybíróság
 - Legfelsőbb Bíróság
 - Döntések Tára
 - Ügyészség
 - Választottbíróság
 - Versenytanács
 - Országos Választási Bizottság
 - Hírközlési Döntvénytár
- Iratmintatár (CompLex Iratmintatár)
- CompLex Döntvénytár
- Nagykommentárok
- CompLex szakadatbázisok
- Bibliográfia
 - Magyar jogi bibliográfia
 - Európai jogi bibliográfia
- Szakmai tartalmak régi nézetben**
 - Adó Jogtár
 - HR Jogtár
 - Közbeszerzés 2011
 - Közjegyző Jogtár
 - Önkormányzat Jogtár
 - Ügyvéd Jogtár

Tárgyszavak keresése

Tárgyszavak száma: 142

Adatkezelés közigazgatási eljárás során
Adatszolgáltatási kötelezettsége ügyfélnek közigazgatási hatósági eljárásban
Alapelvei közigazgatási eljárásnak
Alkotmánybíróság határozata alapján indítható eljárás
Alkotmányjogi panasz
Anyanyelvhasználati joga ügyfélnek közigazgatási eljárás során
Allamigazgatási eljárás; közigazgatási hatósági eljárás és szolgáltatás által
Allamigazgatási eljárás különös szabályai elektronikus ügyintézés során
Beadvány (kérelem); közigazgatási eljárás megindításának szabályai
Belföldi jogsegély

Hivatkozások száma: 15

8/2009. (VII. 3.) IRM utasítás a közigazgatási hatósági eljárásban alkalmazott
4/2009. (VIII. 12.) FVM utasítás a földművelésügyi és vidékfejlesztési miniszte
89/2009. (IX. 24.) AB határozat jogszabály alkotmányellenességének utólagos
83/2009. (IX. 3.) AB határozat jogszabály alkotmányellenességének megállap

Keresés
Szöveg tárgyszavai
Összes tárgyszó

Ugrás

Mégse

2004. évi CXL. törvény
a közigazgatási hatósági eljárás és szolgáltatás általános szabályairól
- I. Fejezet
Alapelvek és alapvető rendelkezések
- Eljárás alapelvek
sználat
y hatálya
- lás
elés
t
g, hatáskör, illetékesség
ság
kör megállapítása és az eljárási kötelezettség
sség
óság, a hatáskör és az illetékesség vizsgálata
ri, illetékességi vita
az illetékességi területen kívül
jogsegély
közi jogsegély
zet
attartás általános szabályai
attartás elektronikus úton
tatás
t
cú eljárás
ás megindítása
- Ideiglenes biztosítási intézkedés
A kérelem érdemi vizsgálat nélküli elutasítása
Az eljárás megszüntetése
Az eljárás felfüggesztése
Ügyintézési határidő

Találati listák

- Ha **hagyományos módszerekkel** navigálunk („browsing”), akkor végül mindig eltalálunk a dokumentumig.
- A keresés jellegzetessége, hogy a megadott kritériumoknak egyszerre **többféle dokumentum is megfelelhet** (sőt tipikusan ez a helyzet) .
- Ekkor a keresés és a dokumentum elolvasása közé beékelődik a **találati lista**.
- A találati listáknak két fajtája van:
 - **Egyszerű, nem rangsorolt lista:** ilyenkor a rendszer a belső felépítése alapján a belső csoportosításait alapul véve listázza ki a megtalált dokumentumokat.
 - **Rangsorolt lista:** valamilyen „fontosság” szerint rangsorolja a dokumentumokat a rendszer.

A relevancia

- **A relevancia** a kereső információ-igénye és a megtalált dokumentum, szövegrészlet, „tétel” közötti távolság mérésére szolgál.
- **Teljesen releváns** egy tétel, ha az információ-igényt teljesen kielégíti, és **teljesen irreleváns**, ha teljesen másra vonatkozik, „köze sincs” a feltett kérdéshez.



Mitől függ a jó találat?

A relevancia, azaz a találat minősége több tényezőn múlik – de
→ **alapvetően a kereső személy „szakmai előismeretein”!** ←

- **Adatbázis-oldali tényezők**
 - Az adat minősége
 - A kereső minősége
(index és kereső-felületek)
- **A kereső személyben rejlő tényezők**
 - A feltett keresőkérdések minősége
 - A keresési stratégia és taktika



A relevancia típusai

- **0/1 típusú relevancia – célzott keresés:**
 - ha egy olyan tételt keresünk, amelynek tudjuk egy vagy két azonosítóját – egy keresett dokumentum
 - Ez a tétel lehet maga a dokumentum, vagy a dokumentumban előforduló számszerű adat, szó, stb. (Pl. keresem a Ket. 12. §-át, vagy mennyi a minimálbér összege) – egy keresett adat
- **Fuzzy relevancia – többé-kevésbé releváns találat:**
 - lehet azért, mert a probléma maga összetett
 - vagy a szöveg homályos
 - Vagy mindkettő, illetve egymásra vonatkoztatottan azok

Az elektronikus forráskutatás fázisai

1. A keresési stratégia, és a keresőkérdések kiválasztása
 1. A jogi probléma megragadása
 2. A keresési stratégia – célzott, vagy nyílt végű keresés
 3. A keresőkérdés kiválasztása – lehetőségek
2. Keresés
3. A találatok értékelése, a jogszabályok értelmezése a másodlagos jogi források segítségével
 1. Döntvénytárak
 2. Kommentárok

Keresési lehetőségek

A dokumentumoknak más jellemzőik is vannak pl. az azokat leíró paraméterek, (cím, típus, keletkezési idő) > **parametrikus keresést** tesz lehetővé – ez a keresés nem a dokumentumban keres.

Egy szavas keresések finomításának lehetőségei

- **Szinonima-szótárak** = emberölésre a gyilkosságot is megtalálja
- **Taxonómiák** = autó szóra az alája sorolt fogalmakat is megtalálja = Ford, Toyota.
- **Pozíció szabályok**: = Nem mindegy, hogy hol fordul elő az adott szó. (Pl. címben, alcímben fontosabb) > (v. mezőben történő keresés)

Több szóból álló kombinációk finomításának lehetőségei

- Még több logikai **operátor**: pl XOR = a kettő közül csak az egyik
 - **Közelségi szabály**: (NEAR) Ha több szóra keresek, akkor nem mindegy ezek a szavak milyen távolságra vannak egymástól > (alapesete a „pontos kifejezés” = 1 szó távolságra helyezkedjen el); de lehet **legalább** 2 szó távolság (ekkor lehet kisebbre)
 - **Sorrendhelyességi szabály** (BEFORE, AFTER) = több szó esetén a sorrend szabály megmondja melyik lehet elől.
- ➔ Mindezek egymással is **kombinálhatóak**, és ezzel nagyon pontos keresések folytathatóak le. Egy idő után a komplikált szabályok már akár a „megértést” is imitálhatják. (Mintha a gép értené a kérdést.)

Tippek

- Gyakori és ritka kifejezések
- Hétköznapi és jogi műszavak
- Iteratív módszer
- A rendszer alapbeállításainak figyelembe vétele
- A hatályosság figyelembe vétele

