

**MTA SZTAKI DSD**

Department  
of Distributed  
Systems

## **Plágiumkeresés különböző nyelvek között**

*Pataki Máté*

- n Bemutató
- n KOPI Plágiumkereső
- n Plágiumkereső működése
- n Plágiumkeresés nyelvek között

- n MTA SZTAKI Elosztott Rendszerek Osztály
- n 1995. óta létezik
- n 10 teljes állású munkatárs, és diákok
- n Kutatás, fejlesztés, (online) szolgáltatások
- n Munkák eloszlása:
  - n 80% EU-s k+f pályázatok
  - n 15% Hazai pályázatok
  - n 5% Belső projektek és szolgáltatások
- n Három fő terület:
  - n Digitális könyvtárak és archívumok
  - n Csoportmunkát támogató technológiák
  - n Webes rendszerek

# World Wide Web

Government Portals  
Infrawebs

Brein E-VOTING  
E-ADMINISTRATION  
Web4Groups

Workflow Promóció  
Forum  
Collaborative Filtering

# Csoportmunka

SZTAKI Szótár  
KOPI

GeneSyS  
StreamOnTheFly  
EUTIST-AMI  
Abilities

CORES

Select Rating

# Digitális Könyvtárak

AQUA

HEKTÁR

DELOS NoE 1

DELOS

ORG

DELOS NoE 2

## További információk

<http://dsd.sztaki.hu>

- n A plágium probléma a
  - n Felsőoktatás területén
  - n Középiskolában is egyre inkább
  - n Tudományos életben
  - n Digitális könyvtárak számára
  - n Könyvkiadóknak
  - n Cégek esetében is (pl. honlapok tartalma)
  - n Wikipedia

## Probléma

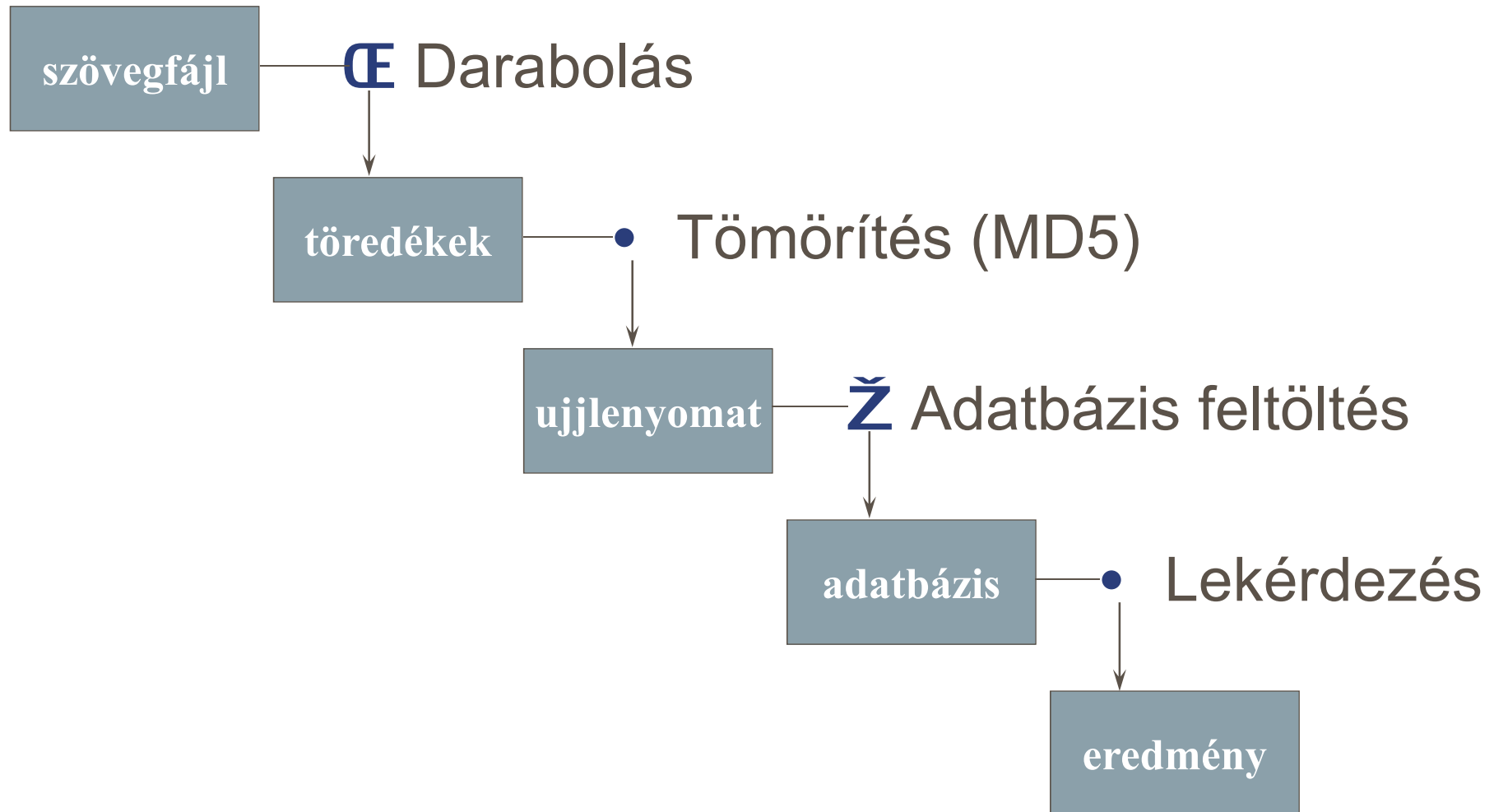
- n Túl sok diploma, cikk, dokumentum, internetes oldal, senki se ismerheti mindnek a tartalmát
- n Nem elég érezni, hogy az adott mű plágium, be is kell bizonyítani
- n Védekezés
  - n Másolás megakadályozása, másolásvédelem
  - n Másolás felderítése, plágiumkeresés

## Plágiumkereső rendszerek

- n Vízjel vagy ellenőrzőösszeg (PDF)
- n Írás stílusa alapján azonosítja a szerzőt (authorship attribution, JGAAP)
- n Nyílt keresőszolgáltatások (pl. google)
- n Szöveges összehasonlítás (MS Word)
- n Tesztet generál a dolgozatból (Glatt)
- n Ismeretlen működésű keresők
  
- n Szövegdarabolással működő rendszerek (KOPI)



## Szövegdarabolással működő rendszerek



# A KOPI Plágiumkereső ismertetése

- n KOPI Online Plágiumkereső és Információs Portál - internetes **hasonlóság** és **plágiumkereső** szolgáltatás
- n Partner: Monash University, Melbourne
- n 2002 ITEM pályázat (IHM-OM)
- n MTA-SZTAKI Elosztott Rendszerek Osztály
- n <http://kopi.sztaki.hu/>

sztaki kopi Kopi Online Plágiumkereső és Információs Portál

Szótár KOPI NDA Kereső

Plágiumkeresés és dokumentumkezelés

Feltöltés Dokumentumaim Plágiumkereső Futó keresések

Válassza ki azokat a dokumentumokat amelyekkel plágiumkeresést szeretne végezni.

Cím	Szerző	Feltöltés dátuma	Szűrő bekapcsolása
<input checked="" type="checkbox"/> Cikk 3 PZsP 01b 2% (30 szó) egyezés	-	2005.09.30.	<input type="checkbox"/> Szerkeszt <input type="checkbox"/> Részletes
<input checked="" type="checkbox"/> Szabványok a kórházi informatikában - Absztrakt 30% (30 szó) egyezés	-	2005.09.30.	<input type="checkbox"/> Szerkeszt <input type="checkbox"/> Részletes
<input checked="" type="checkbox"/> A kutatók kötelesek	IP	2005.09.13.	<input type="checkbox"/> Szerkeszt <input type="checkbox"/> Részletes

magyar | english  
Betűméret - +  
Nagy kontraszt  
Súgó

KOPI  
Kezdőlap  
Fórum

Felhasználó: a  
Beállításaim  
Üzenetek  
Plágiumkeresés  
Kilépés  
Admin

- n A KOPI Rendszerbe feltöltött dokumentumok
  - n A plagizálás gyorsan kiderül
  - n A forrás(oka)t is megjelöli
  - n Kisebb – akár egy-két mondatnyi – rész egyezőségét is kijelzi
  - n Megkerülése nehézkes és nem automatizálható
  - n A lebukás kockázata jelentősen megnő
  - n Másolásvédelemmel együtt is használható

- n A mű szabadon terjeszthető
  - n Nem kell másolásvédelemmel foglalkozni
  - n A keresőrendszerek is hozzáférnek (google)
  - n Több ember olvashatja
  - n Több ember fog rá hivatkozni
- n Jelenlegi rendszer hátrányai
  - n egy egységes rendszert érdemes használnia mindenkinek
  - n fel kell tölteni a védeni kívánt dokumentumokat a rendszerbe, vagy elérhetővé kell tenni

- n Sok esetben nem szándékos a plagizálás
- n Nem oktatják az egyetemeken a helyes idézés módját
- n Mekkora hasonlóságot várunk el
  - n 0% - nincs irodalomkutatás
  - n 10%
  - n 50%
  - n 100% - egyértelműen plágium
- n Diákok és tanárok egyaránt használják a KOPI Plágiumkeresőt

- n Több mint 13 000 felhasználó
- n 20 000 dokumentum
- n Közel 20 000 000 dokumentumrészlet
  - n Ebből 13 000 000 magyar
- n BME, 400-500 diák, 5 feladat, 6 év
  - n 2007: 9 pár, 2009: 4 pár, 2010: 2 pár

```
545: <!-- A properties pedig tartalmazza az összes többi adatot az utasnak -->
```

```
546: <utas>
```

```
547: <xsl:for-each select="page/ROWSET/ROW">
```

```
548:     <name><xsl:value-of select="NEV"/></name>
```

```
549:     <properties>
```

```
550:
```

## Vélemény a mérésről

A mérés számomra egész érdekesnek bizonyult, annak ellenére, hogy sok időt vett igénybe. Hasznos dolognak tartom a kategóriákon belüli kereséseket.

```
557: </
```

```
558: </
```

```
559: </
```

```
560: </
```

```
561:
```

```
562: A szöveg.xml fájlomban egy sorba iratom ki egy utas jellemzőit pontosvesszővel elválasztva. Ha több utast iratok ki, akkor minden utas új sorba kerül. Itt is meg kell adnom az
```

```
563:
```

```
564: Vélemény a mérésről
```

```
565: A mérés számomra egész érdekesnek bizonyult, annak ellenére, hogy sok időt vett igénybe. Hasznos dolognak tartom a kategóriákon belüli kereséseket.
```

```
566:
```

```
567:
```

```
422: <utas>
```

```
423: <xsl:for-each select="page/ROWSET/ROW">
```

```
424:     <name><xsl:value-of select="NEV"/></name>
```

```
425:     <properties>
```

```
426:
```

## Vélemény a mérésről

A mérés számomra egész érdekesnek bizonyult, annak ellenére hogy sok időt vett igénybe. Hasznos dolognak tartom a kategóriákon belüli kereséseket és a dinamikus oldalmegvalósításokat.

```
430:
```

```
431: </
```

```
432: </
```

```
433: </
```

```
434: </
```

```
435:
```

```
436: A
```

```
437:
```

```
438: A teljes mérés a ...
```

```
439:
```

```
440: Vélemény a mérésről
```

```
441: A mérés számomra egész érdekesnek bizonyult, annak ellenére hogy sok időt vett igénybe. Hasznos dolognak tartom a kategóriákon belüli kereséseket és a dinamikus oldalmegvalósít
```

```
442:
```

```
443:
```

## Plágiumkeresés nyelvek között

- n A projekt célja olyan algoritmusok elkészítése, amelyek lehetővé teszik,
  - n az angol eredeti szöveg megtalálását a magyar fordítás ismeretében
  - n interneten vagy digitális könyvtárakban, gyűjteményekben.
  
- n Felhasználási területek:
  - n plágiumok felkutatása
  - n párhuzamos korpuszok automatikus építése
  - n hírek, cikkek, anyagok terjedésének a vizsgálata a hálózaton
  - n hasonló témákkal dolgozó emberek, kutatócsoportok felkutatása



## Téma háttere

- n Európában fontos téma
- n Legaktívabbak: németek, spanyolok, csehek
- n Az algoritmusok nyelvpár-függők
  - n pl.: angol, német, spanyol, francia, holland nyelvpárnál jó
  - n angol-lengyel esetében használhatatlan
- n Magyar nyelvben három főbb hátrány
  - n nem kötött szórend
  - n ragozás
  - n jelentős nyelvtani különbség az angol nyelvtől
- n rosszak az automatikus fordítók (erre)

- n Test cases for plagiarism detection software, Debora Weber-Wulff, HTW Berlin, 2010
- n 48 különböző plágiumkereső, 42 teszt
- n *The biggest gap in all the plagiarism checkers was the inability to locate translated plagiarism. While this is widely expected as the technology to make such detections simply is not there.*

- n Gépi fordítás
  - n Google translate
    - n Hogy lássuk ezt meg kell magyaráznunk az anarchizmus alapjául szolgáló morális szemléletmódot
    - n To see this of course we must expound the moral outlook underlying anarchism
    - n To see this you have to explain the underlying moral approach to anarchism
    - n Hogy látja ezt persze meg kell kifejtteni a morális kilátások alapjául szolgáló anarchizmus

**HUN: A Carnatic ez? - robbant ki.**

**ENG: Am I on the Carnatic?" -1**

**HUN: A detektívnek minden oka megvolt, hogy így okoskodjon.**

**ENG: The detective was not far wrong in making this conjecture. -2**

**HUN: A detektív is hasztalanul fáradozott, hogy ő legyen a nézeteltérésben a főszereplő.**

**ENG: As vainly did the detective endeavor to make the quarrel his. -3**

**HUN: - A hídon.**

**ENG: "On the bridge." 2**

**HUN: - Addig óvadék ellenében szabadlábra helyezem mindkettőjüket.**

**ENG: "Meanwhile, you are liberated on bail." -3**

**HUN: A gentlemannek különben is kész volt a terve a továbbiakra.**

**ENG: Mr. Fogg's course, however, was fully decided upon. -6**

**HUN: A gépész azonban történetesen épp e napon felment a fedélzetre, megkereste Mr. Foggot, és meglehetősen élénk vitát folytatott vele.**

**ENG: On this day the engineer came on deck, went up to Mr. Fogg, and began to speak earnestly with him. -4**

# Magyar mondat fordításának a keresése angol oldalon

Mondat:

A teljes lista megtekintéséhez és referenciáink leírásához nézzon bele ebbe a szekcióba.

URL:

<http://dsd.sztaki.hu/en>

Keres

16 sentences found on <http://dsd.sztaki.hu/en>.

One similar sentence found:

- For a complete list and description of our references, look into this section. (2)

## Magyar mondat fordításának a keresése angol oldalon

Magyar URL:  
  
 Angol URL:

66 sentences found on <http://dsd.sztaki.hu/mission/hu> and 59 sentences found on <http://dsd.sztaki.hu/mission/en>.

HUN: MTA SZTAKI Elosztott Rendszerek Osztály  
 ENG: MTA SZTAKI - Department of Distributed Systems (2)

HUN: A DSD jelenlegi kulcsszavai SOA, szolgáltatás orientált architektúrák, szemantikus web szolgáltatások, ontológiák, proaktív projekt menedzsment, awareness, p2p, peer-to-peer, over  
 alkalmazások, metaadat séma regiszterek, web 2.0, gazdag felhasználói felületek, AJAX, plágium, plágiumkeresés, LLD (Library Linked Data), LOD (Linked Open Data), RDF, szókinccs, S  
 ENG: Actual keywords of the Department SOA, service oriented architectures, context awareness in collaborative systems, semantic web services, ontologies, proactive project management  
 mobile applications, metadata registry, web 2.0, rich application interfaces, AJAX, plagiarism, LLD (Library Linked Data), LOD (Linked Open Data), RDF, vocabulary, SKOS, accessibility

HUN: S-CUBE: European Network of Excellence in Software Services and Systems  
 ENG: S-CUBE: European Network of Excellence in Software Services and Systems (9)

HUN: CrossMedia: Elosztott szemantikus adatok hálójának szolgáltatása (ASP) rendszer  
 ENG: CrossMedia: ASP solution for advanced distributed semantic processing (1)

HUN: W3C-HU: World Wide Web  
 ENG: W3C-HU: World Wide Web (1)

HUN: BREIN: Üzleti célvezérelt, megbízható és intelligens Gridek  
 ENG: BREIN: Business objective driven reliable and intelligent Grids for real business (4)

HUN: T-ONE: Automatic Metadata Extraction and User Modeling in Web Portals  
 ENG: T-ONE: Automatic Metadata Extraction and User Modeling in Web Portals (12)

HUN: DELOS  
 ENG: DELOS

HUN: INFRAWEBs: Intelligent Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled Applications Using Semantic Web Technologies, Distributed  
 ENG: INFRAWEBs: Intelligent Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled Applications Using Semantic Web Technologies, Distributed

HUN: ORG: Országos Rákregiszter GRID  
 ENG: ORG: National Cancer Registry GRID (4)

HUN: T-SOA: Transaction and Handling of Concurrency in Service Oriented Architectures  
 ENG: T-SOA: Transaction and Handling of Concurrency in Service Oriented Architectures (11)

HUN: GeneSys: Generikus rendszerfelügyeleti rendszer  
 ENG: GeneSys: Generic Systems Supervision (2)

## Magyar és angol szöveg mondatainak párosítása

Magyar szöveg:

Aarfy mindig igyekezett Natelyn segíteni mert Nately apja gazdag és befolyásos ember volt aki kitűnő állása révén segíthetett volna Aarfyn a háború után.

Keres

Angol szöveg:

Aarfy was always trying to help Nately because Nately's father was rich and prominent and in an excellent position to help Aarfy after the war.

Comparing 1 Hungarian sentence with 1 English.

HUN: Aarfy mindig igyekezett Natelyn segíteni mert Nately apja gazdag és befolyásos ember volt aki kitűnő állása révén segíthetett volna Aarfyn a háború után.  
ENG: Aarfy was always trying to help Nately because Nately's father was rich and prominent and in an excellent position to help Aarfy after the war. (9)

One similar sentence found.

## Két angol szöveg mondatainak párosítása

Egyik szöveg:	Másik szöveg:
<p>And God spoke all these words:            "I am the LORD your God, who brought you out of Egypt, out of the land of slavery.            "You shall have no other gods before[a] me.            "You shall not make for yourself an image in the form of anything in heaven above or on the earth beneath or in the waters below. 5 You shall not bow down</p>	<p>Then God spoke all these words, saying,            I am the LORD your God, (B)who brought you out of the land of Egypt, out of the house of slavery.            You shall have no other (D)gods [a]before Me.            You shall not make for yourself [b]an idol, or any likeness of what is in heaven above or on the earth beneath or in the water under the earth.</p>
<input type="text" value="Keres"/>	

Comparing 21 English sentences with 21 English.

TXT1: And God spoke all these words:  
 TXT2: Then God spoke all these words, saying. (6)

TXT1: "I am the LORD your God, who brought you out of Egypt, out of the land of slavery.  
 TXT2: I am the LORD your God, (B)who brought you out of the land of Egypt, out of the house of slavery. (14)

TXT1: "You shall have no other gods before[a] me.  
 TXT2: You shall have no other (D)gods [a]before Me. (11)

TXT1: "You shall not make for yourself an image in the form of anything in heaven above or on the earth beneath or in the waters below. 5 You shall not bow down to them or worship them; for I, the LORD your God, am a jealous God, punishing the children for the sin of the parents to the third and fourth generation of those who hate me, 6 but showing love to a thousand generations of those who love me and keep my commandments.  
 TXT2: You shall not worship them or serve them; for I, the LORD your God, am a (G)jealous God, (H)visiting the iniquity of the fathers on the children, on the third and the fourth generations of those who hate Me, (22)

TXT1: "You shall not misuse the name of the LORD your God, for the LORD will not hold anyone guiltless who misuses his name.  
 TXT2: You shall not take the name of the LORD your God in vain, for the LORD will not leave him unpunished who takes His name in vain. (10)

TXT1: "Remember the Sabbath day by keeping it holy. 9 Six days you shall labor and do all your work, 10 but the seventh day is a sabbath to the LORD your God.  
 TXT2: For in six days the LORD made the heavens and the earth, the sea and all that is in them, and rested on the seventh day, therefore the LORD blessed the sabbath day and made it holy. (9)

TXT1: On it you shall not do any work, neither you, nor your son or daughter, nor your male or female servant, nor your animals, nor any foreigner residing in your towns. 11 For in six days the LORD made the heavens and the earth, the sea, and all that is in them, but he rested on the seventh day.  
 TXT2: For in six days the LORD made the heavens and the earth, the sea and all that is in them, and rested on the seventh day; therefore the LORD blessed the sabbath day and made it holy. (6)

TXT1: Therefore the LORD blessed the Sabbath day and made it holy.  
 TXT2: Remember (K)the sabbath day, to keep it holy. (2)

TXT1: "Honor your father and your mother, so that you may live long in the land the LORD your God is giving you.  
 TXT2: Honor your father and your mother, that your (P)days may be prolonged in the land which the LORD your God gives you. (10)

TXT1: "You shall not murder.  
 TXT2: You shall not murder. (4)

TXT1: "You shall not commit adultery.  
 TXT2: You shall not commit adultery. (6)

TXT1: "You shall not steal.  
 TXT2: You shall not steal. (4)

TXT1: "You shall not give false testimony against your neighbor.  
 TXT2: You shall not bear false witness against your (U)neighbor. (6)

TXT1: "You shall not covet your neighbor's house.  
 TXT2: You shall not bear false witness against your (U)neighbor. (6)

TXT1: You shall not covet your neighbor's wife, or his male or female servant, his ox or donkey, or anything that belongs to your neighbor.  
 TXT2: You shall not covet your neighbor's house; (W)you shall not covet your neighbor's wife or his male servant or his female servant or his ox or his donkey or anything that belongs to your neighbor. (24)



**<http://kopi.sztaki.hu>**

# Köszönöm a figyelmet!

**Web:** <http://dsd.sztaki.hu>

**Email:** [Mate.Pataki@sztaki.hu](mailto:Mate.Pataki@sztaki.hu)