Szabad szoftveres OCR

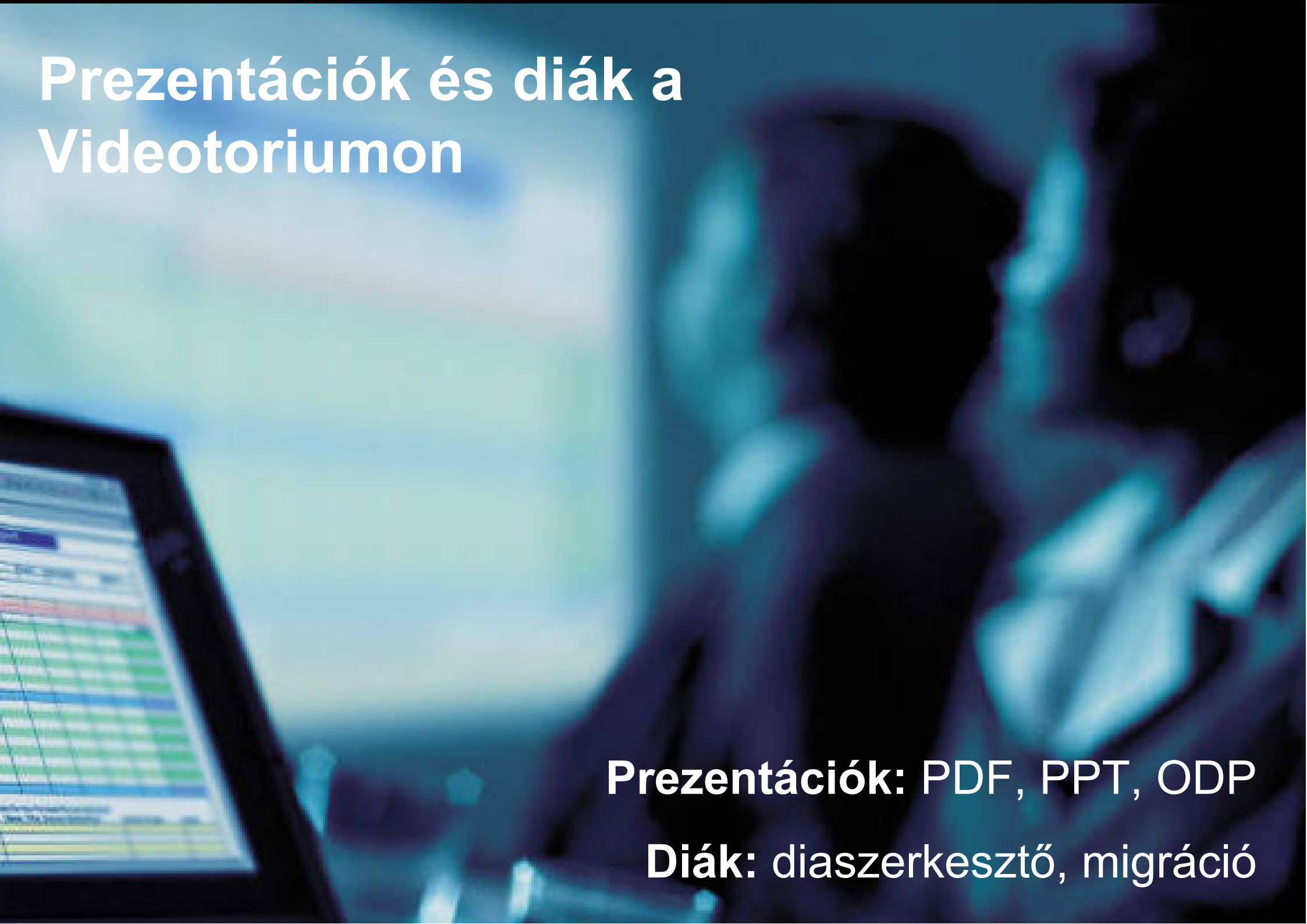
a Videotorium keresőmotor támogatására

Turcsányi Tamás, NIIF

Videotorium 2. workshop - Budapest, 2010. december 19.

Miért kell a prezentáció?

Prezentációk és diák a Videotoriumon



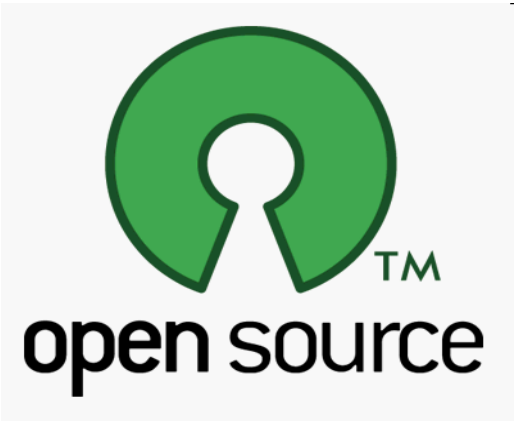
Prezentációk: PDF, PPT, ODP

Diák: diaszerkesztő, migráció

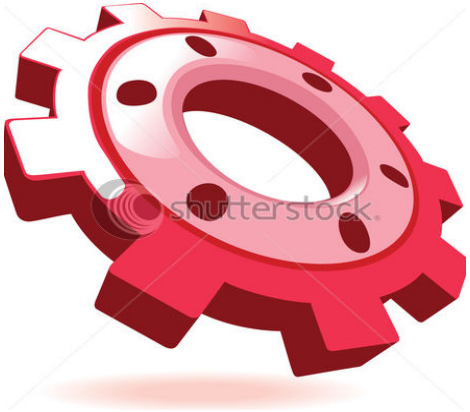

Diák visszakereshetősége: a kezdetek

Célok

Ingyenes, OS megoldás

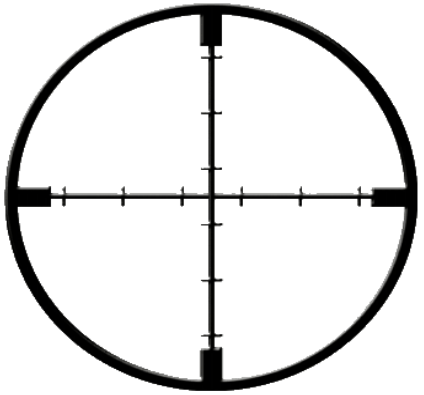


~46000 dia feldolgozása



Bitkép feltöltések

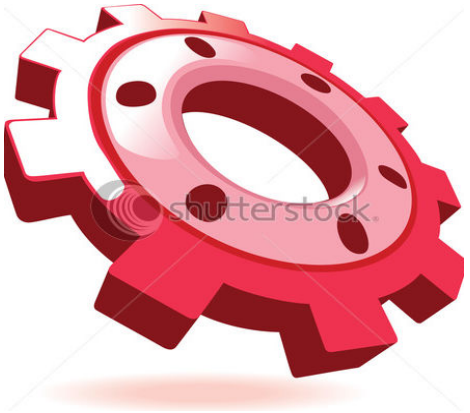
Precíz kereső



Az ideális OCR szoftver

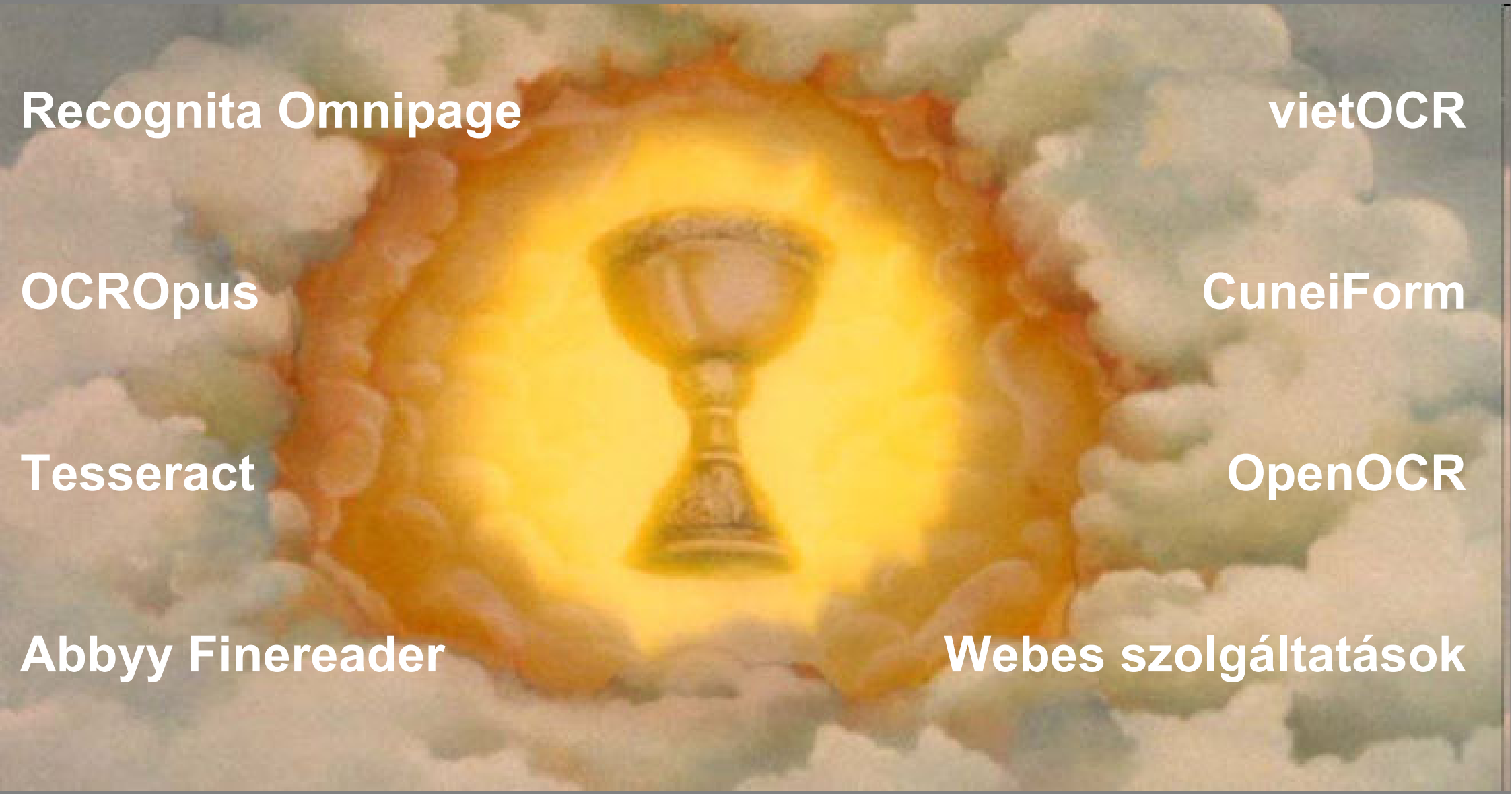
OS

Parancssor

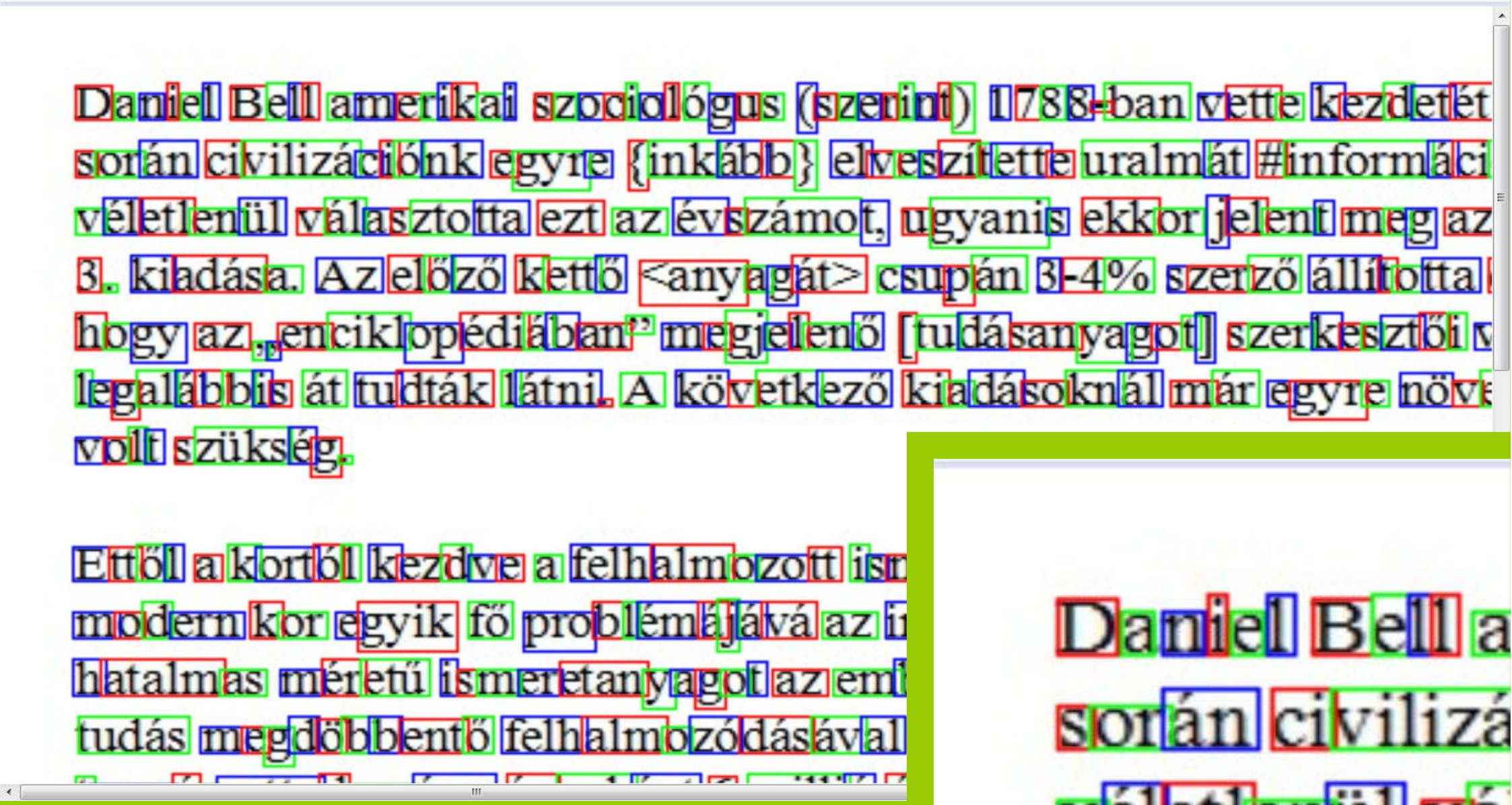


Stabil

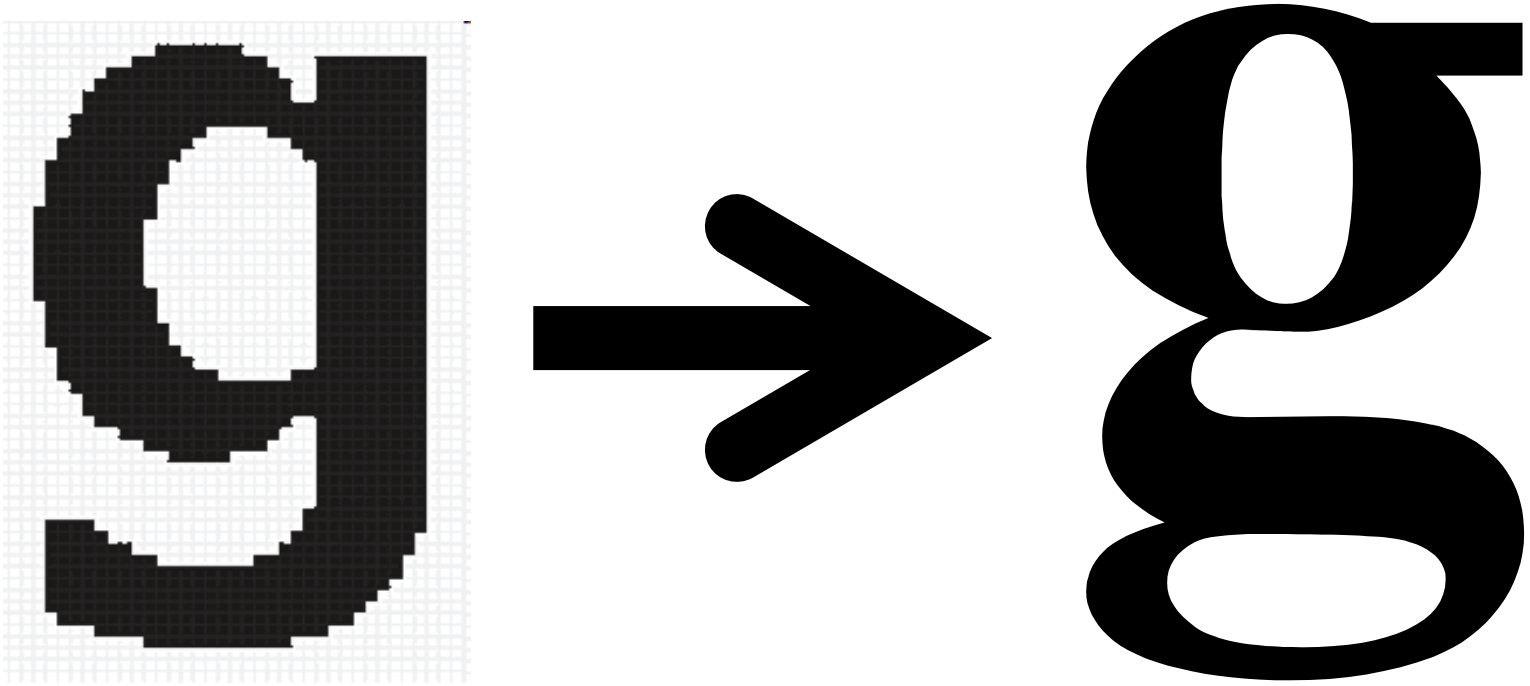
Az ideális OCR szoftver


Recognita Omnipage
vietOCR
OCROpus
CuneiForm
Tesseract
OpenOCR
Abbyy Finereader
Webes szolgáltatások

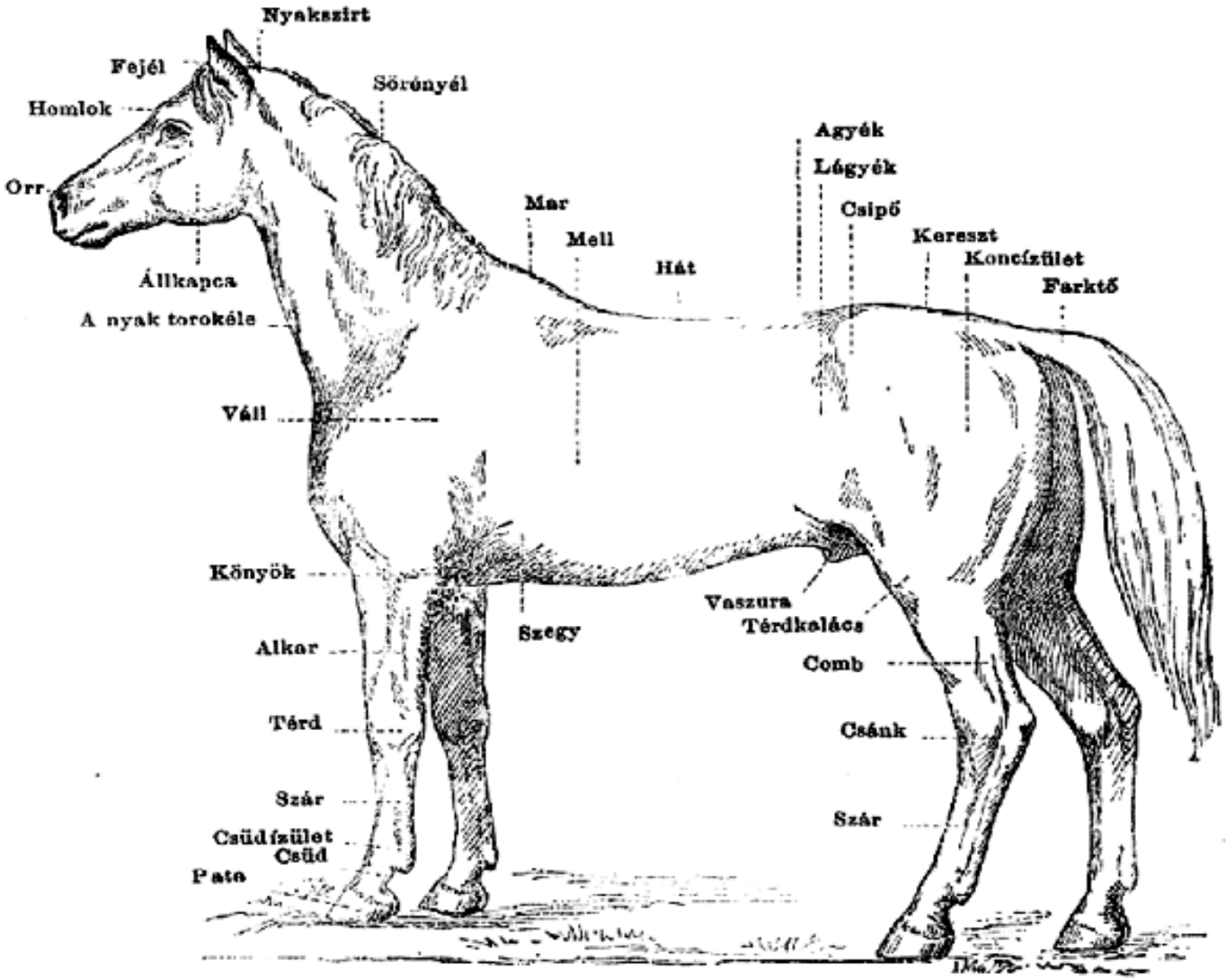
Tesseract tréning



Amit az OCR szeret



Amink van


Nyakszirt
Fejél
Homlok
Orr
Állkapca
A nyak torokéle
Sörényél
Vál
Könyök
Alkar
Térd
Szár
Csüdízület
Csüd
Pata
Szegy
Mar
Mell
Hát
Agyék
Lágyék
Csipő
Kereszt
Koncízület
Farktő
Vaszura
Térdkalács
Comb
Csánk
Szár

Hol tartunk ma?

Tétényi István
NIIF MT

Networkshop 2003

Pécs

ESZKÖZÖK:

INFORMÁCIÓS TÁRSADALOM: E-EUROPE;

INTEGRÁLT ÉS LIBERALIZÁLT KOMMUNIKÁCIÓ;

SZABÁLYOZOTT E-KERESKEDELEM.

ERA: KOORDINÁLT TUDOMÁNYPOLITIKA;

KIVÁLÓSÁGI KÖZPONTOK ÉS HÁLÓZATAIK;

INFRASTRUKTÚRA; 169§; MOBILITÁS;

INTEGRÁLT PROJEKTEK; FP6; NAGYSEBESSÉGŰ

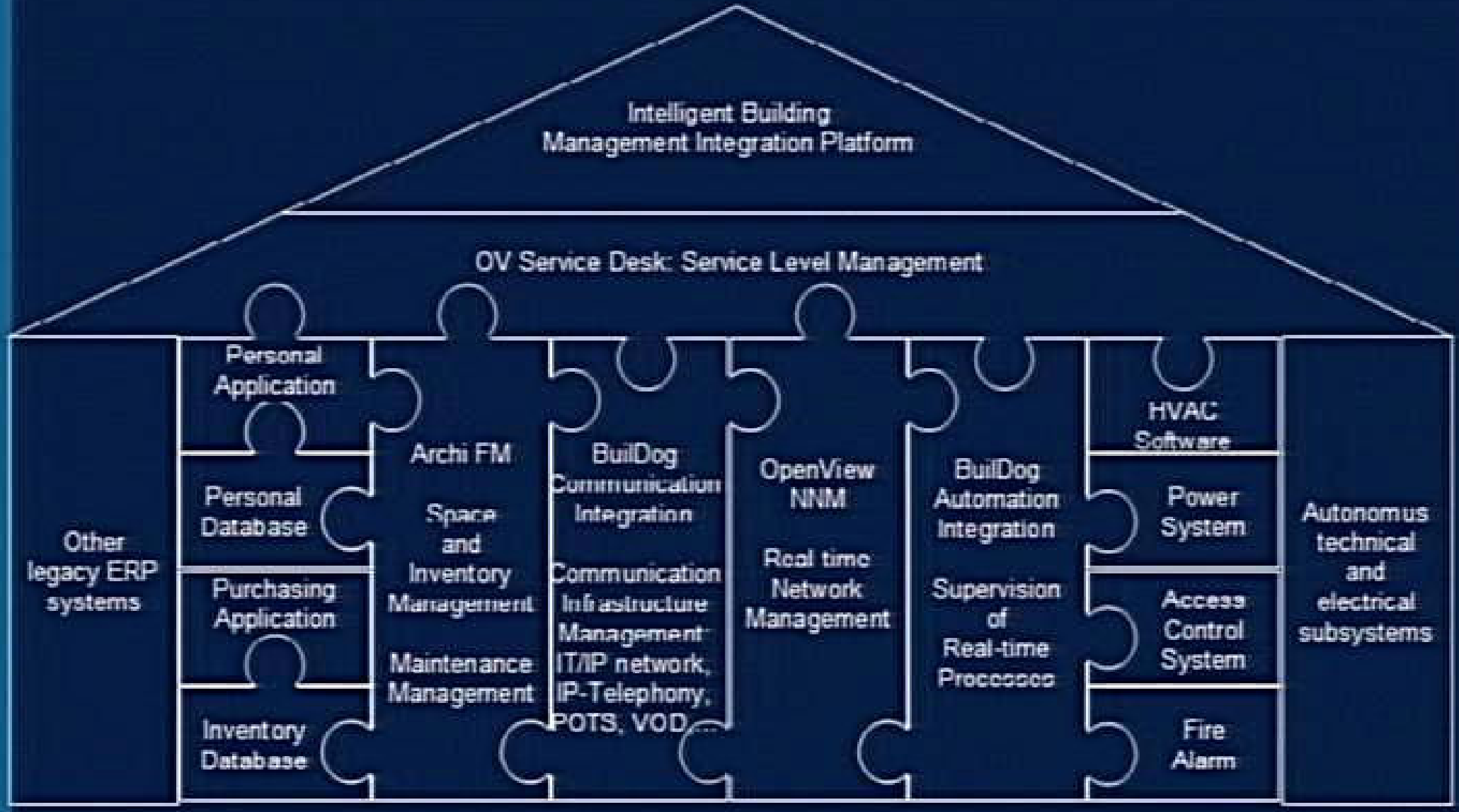
ELEKTRONIKUS HÁLÓZAT.

KÉRDŐJEL: SZAVAK ÉS TETTEK ARÁNYA.

Eszköz-tender I.

- **Közbeszerzési eljárás:**
  - 2003 június – 2003 szeptember
  - 1db nagyteljesítményű Multipoint Control Unit (MCU)
  - 40db professzionális H.323 végberendezés (15 helyett!)
  - Szállító: LNX Hálózatintegrációs Rt.
  - Gyártó: Polycom (USA)
- **MCU:**
  - Accord MGC-100 (Polycom)
  - 16 kártyahely
  - Max. 24 @ 768 Kbps
  - Max. 16 CP konferenciában

Az intelligens épület elemei és kapcsolatai



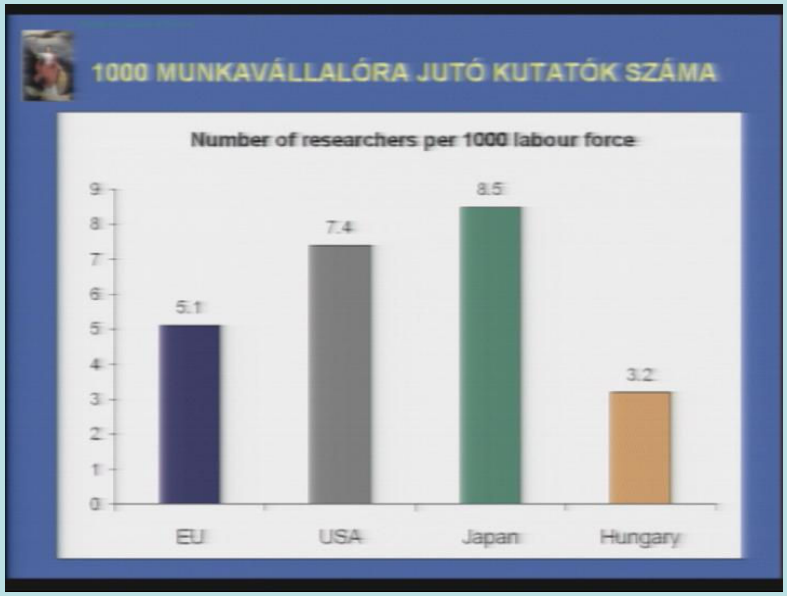
Előfeldolgozás

adaptív

Lehetőségek

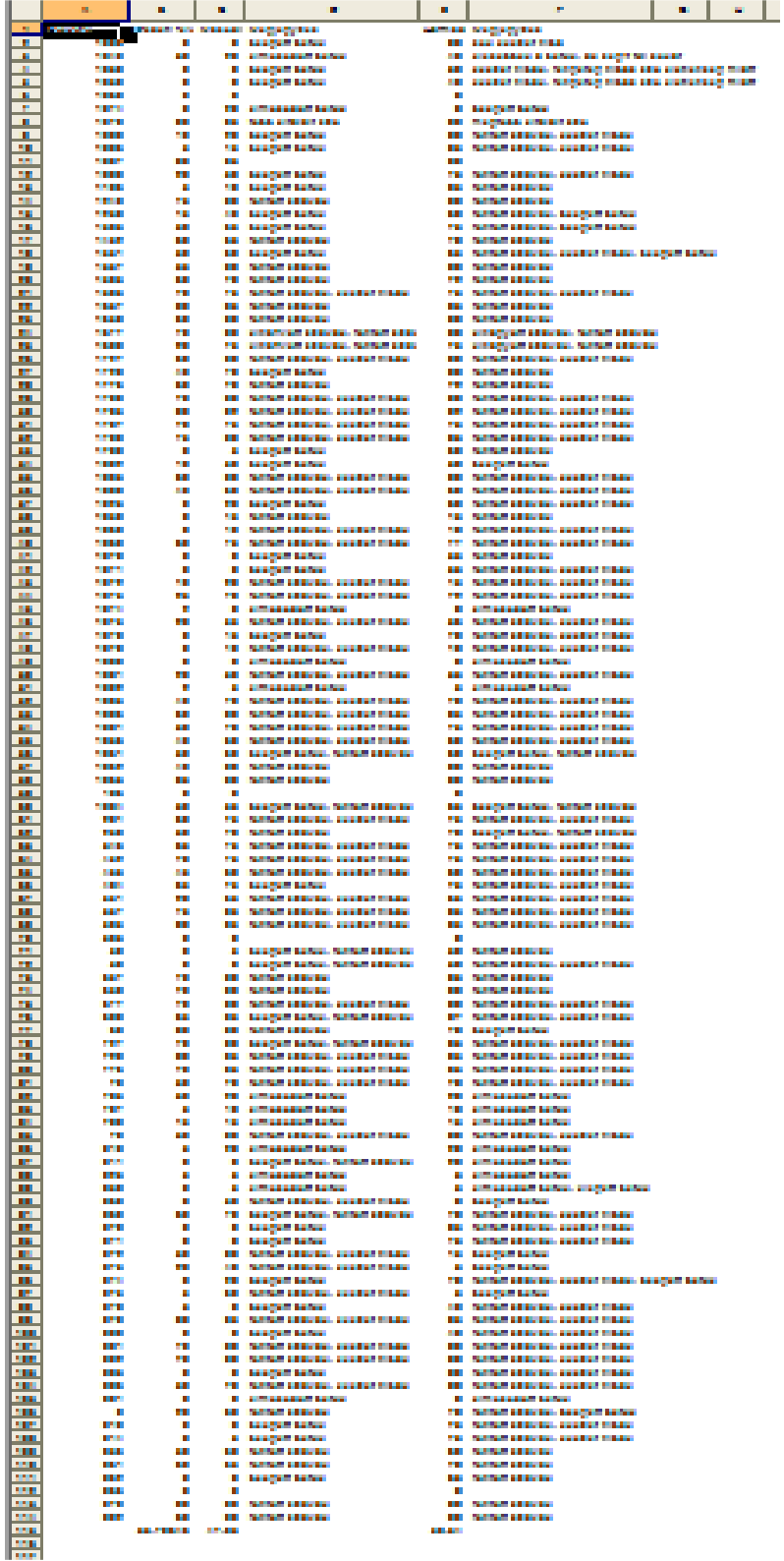
középutas „adaptív”

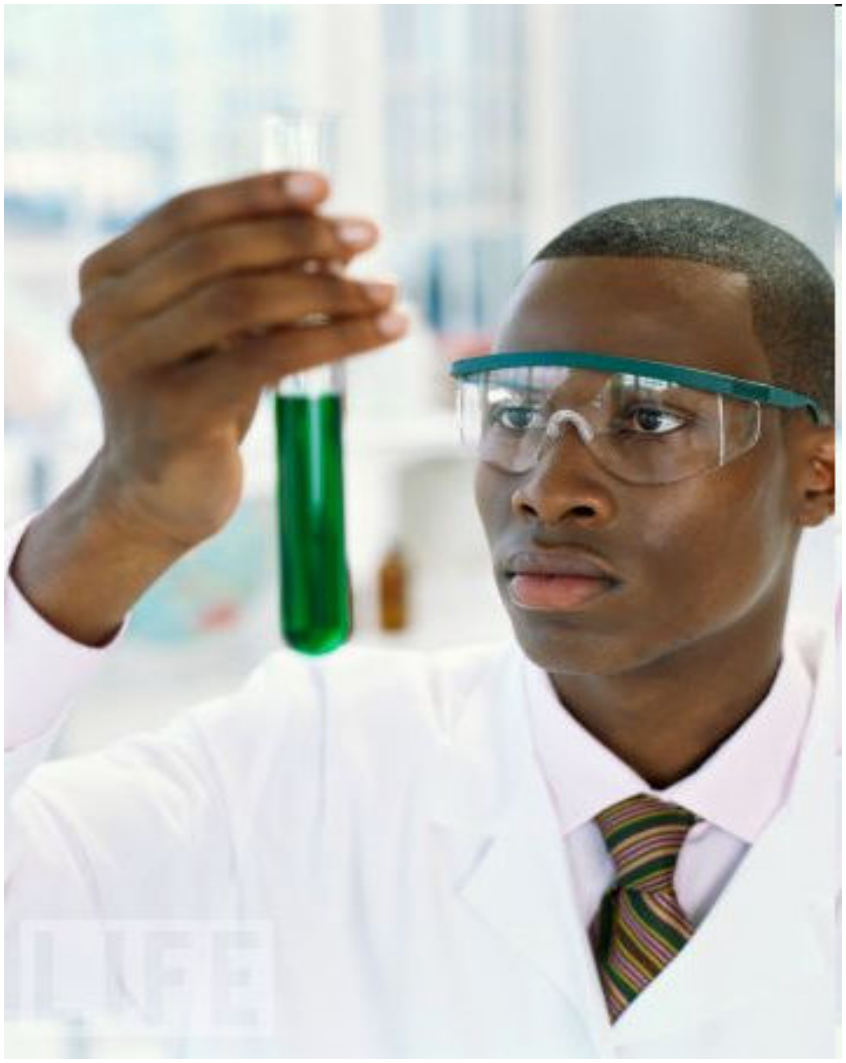
KERESSÜK!



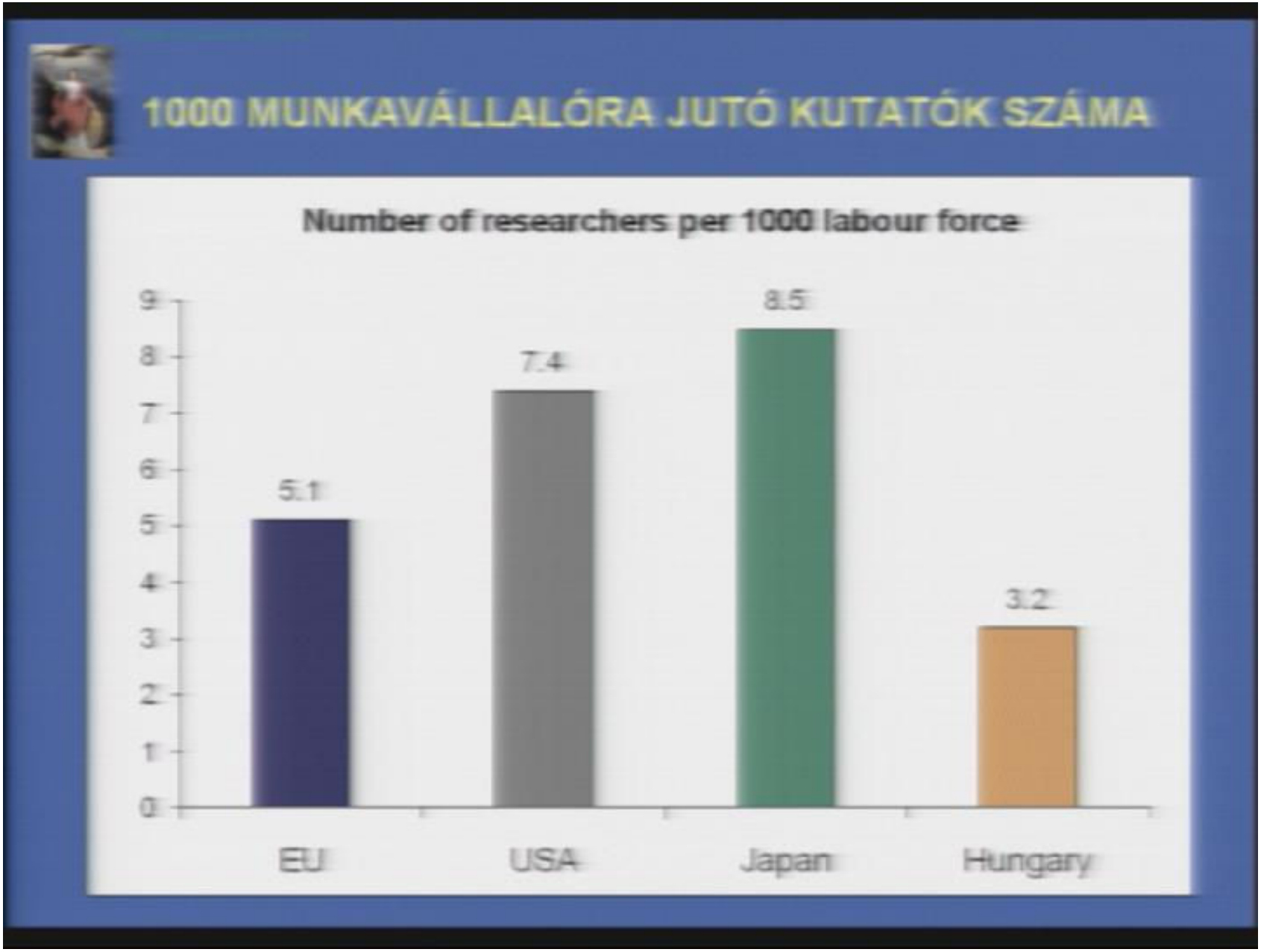
Aki bármit tud a fenti diáról,
kérjük hívja az alábbi számot:

06-55/555-NIIF

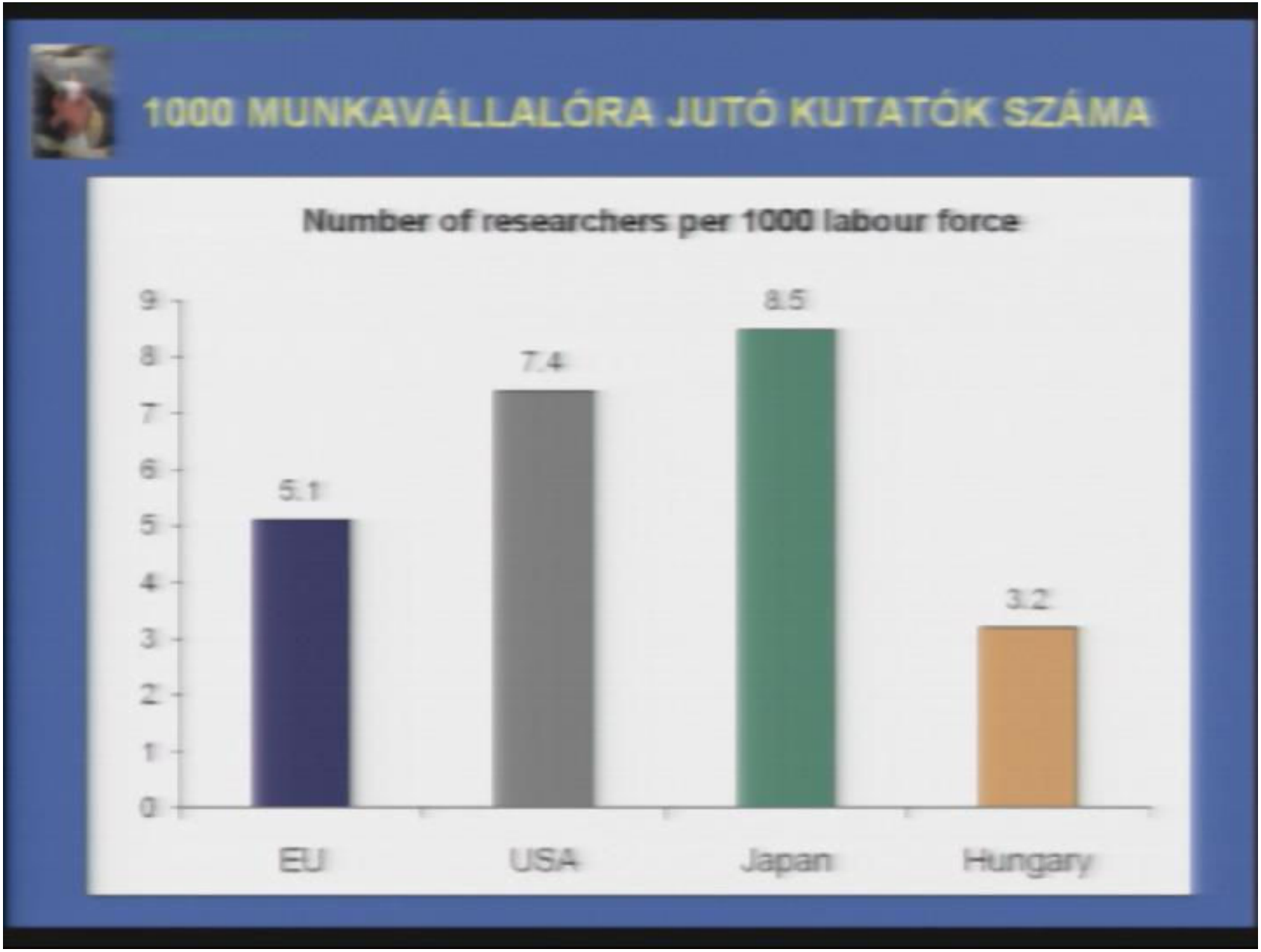




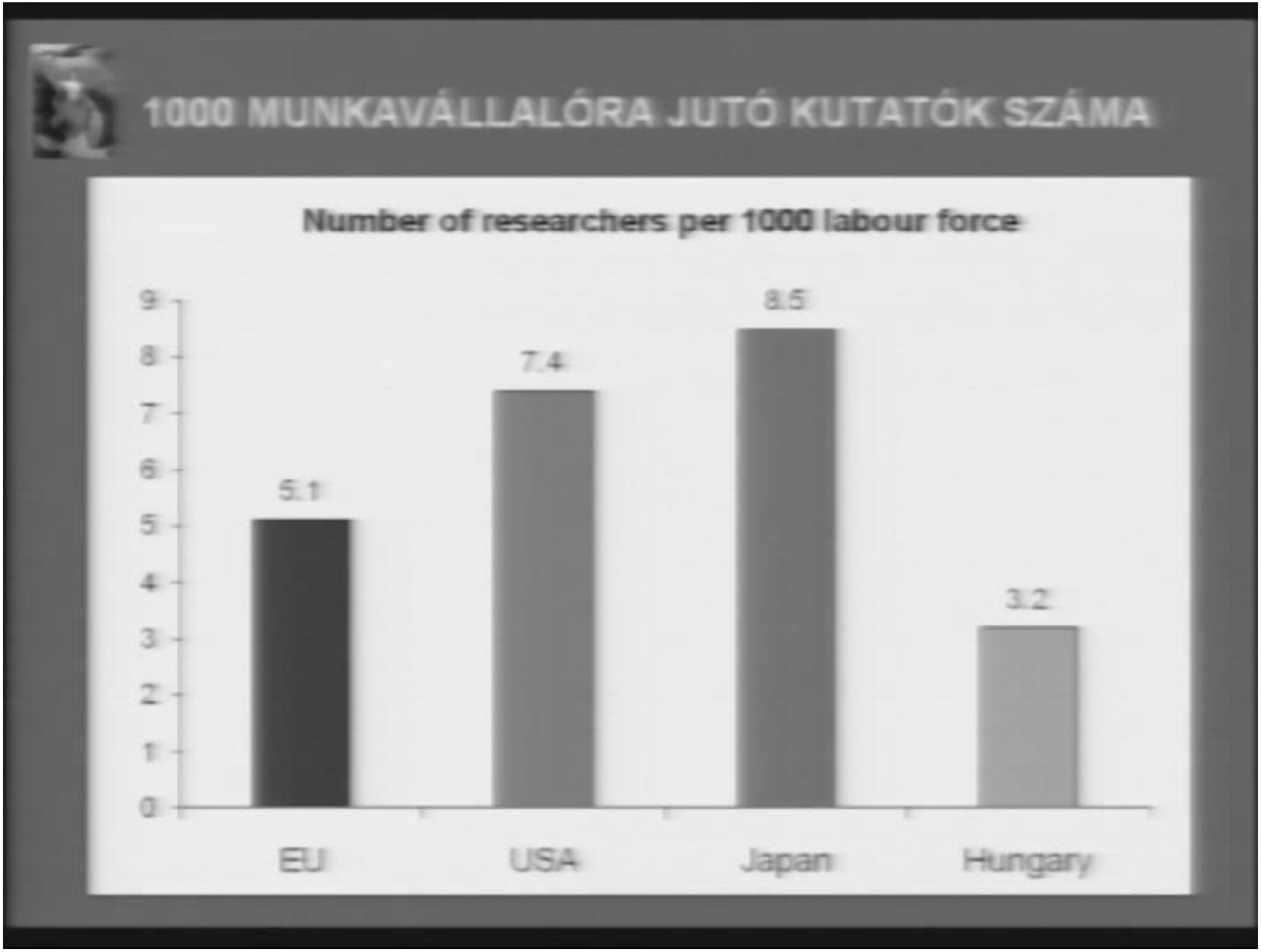
Tesztalany


1000 MUNKAVÁLLALÓRA JUTÓ KUTATÓK SZÁMA
Number of researchers per 1000 labour force
9
8
7
6
5
4
3
2
1
0
5.1
7.4
8.5
3.2
EU
USA
Japan
Hungary

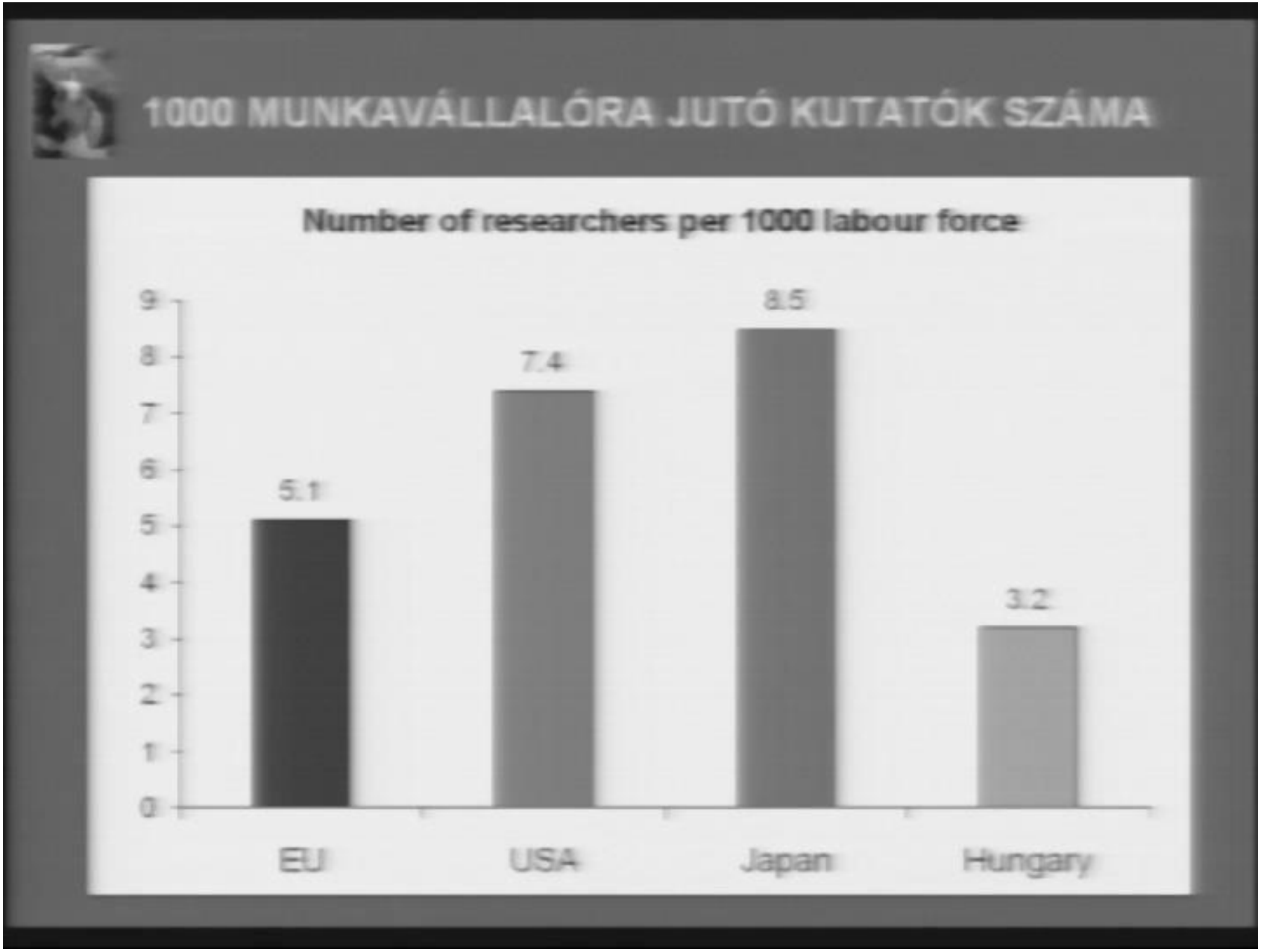
-despeckle


1000 MUNKAVÁLLALÓRA JUTÓ KUTATÓK SZÁMA
Number of researchers per 1000 labour force
9
8
7
6
5
4
3
2
1
0
5.1
7.4
8.5
3.2
EU
USA
Japan
Hungary

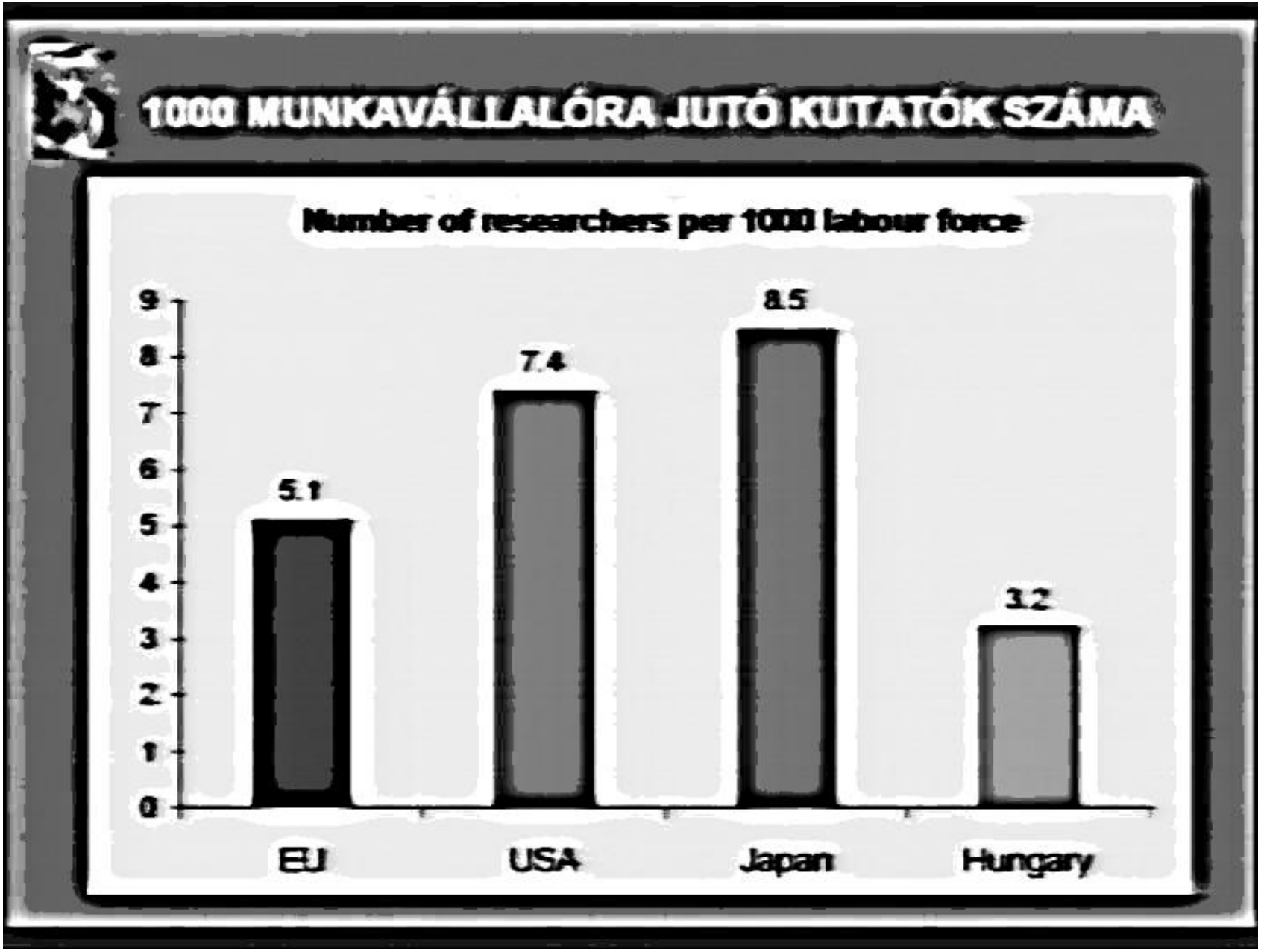
-colorspace gray



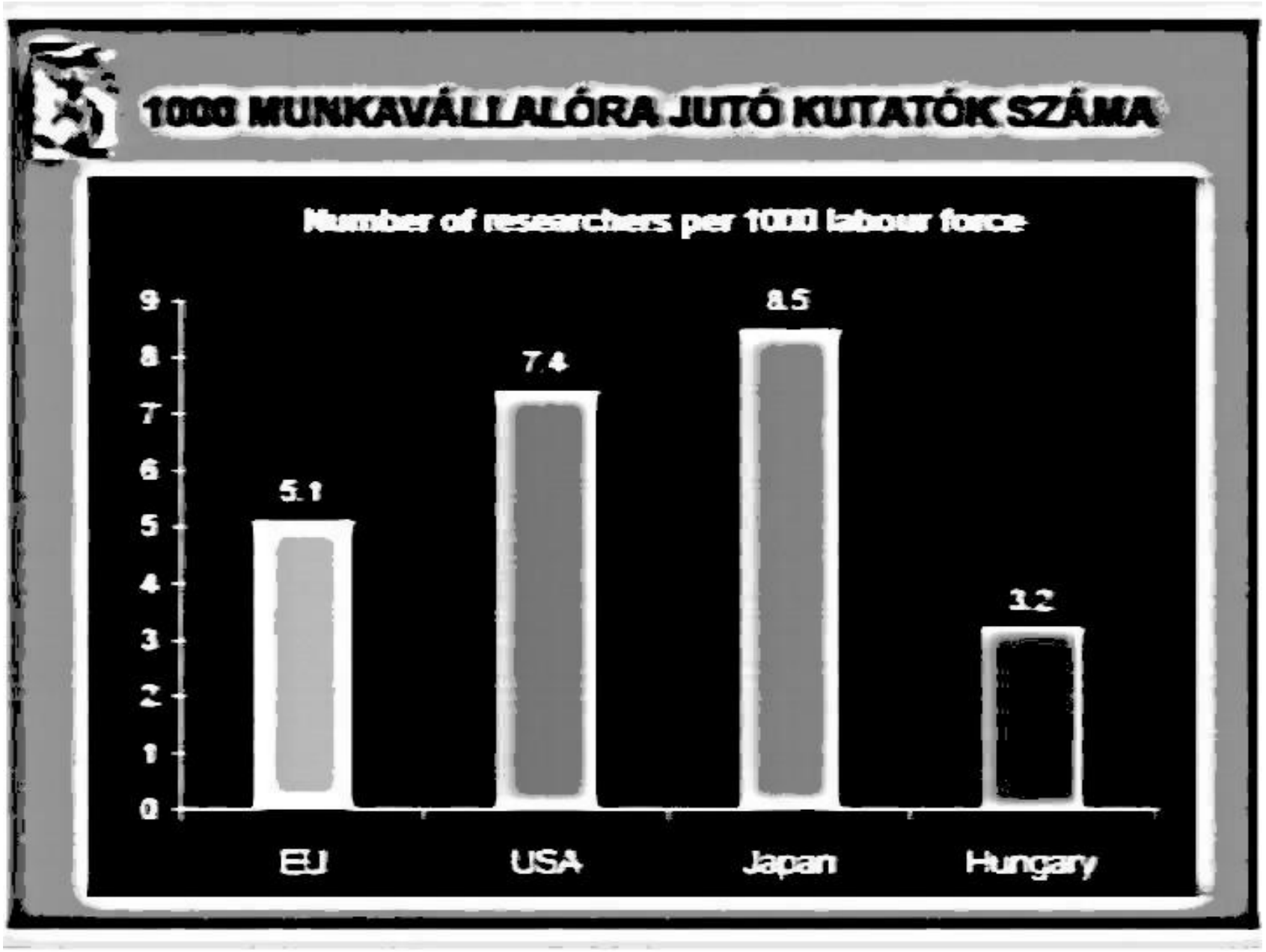
-resize 200%


1000 MUNKAVÁLLALÓRA JUTÓ KUTATÓK SZÁMA
Number of researchers per 1000 labour force
9
8
7
6
5
4
3
2
1
0
5.1
7.4
8.5
3.2
EU
USA
Japan
Hungary

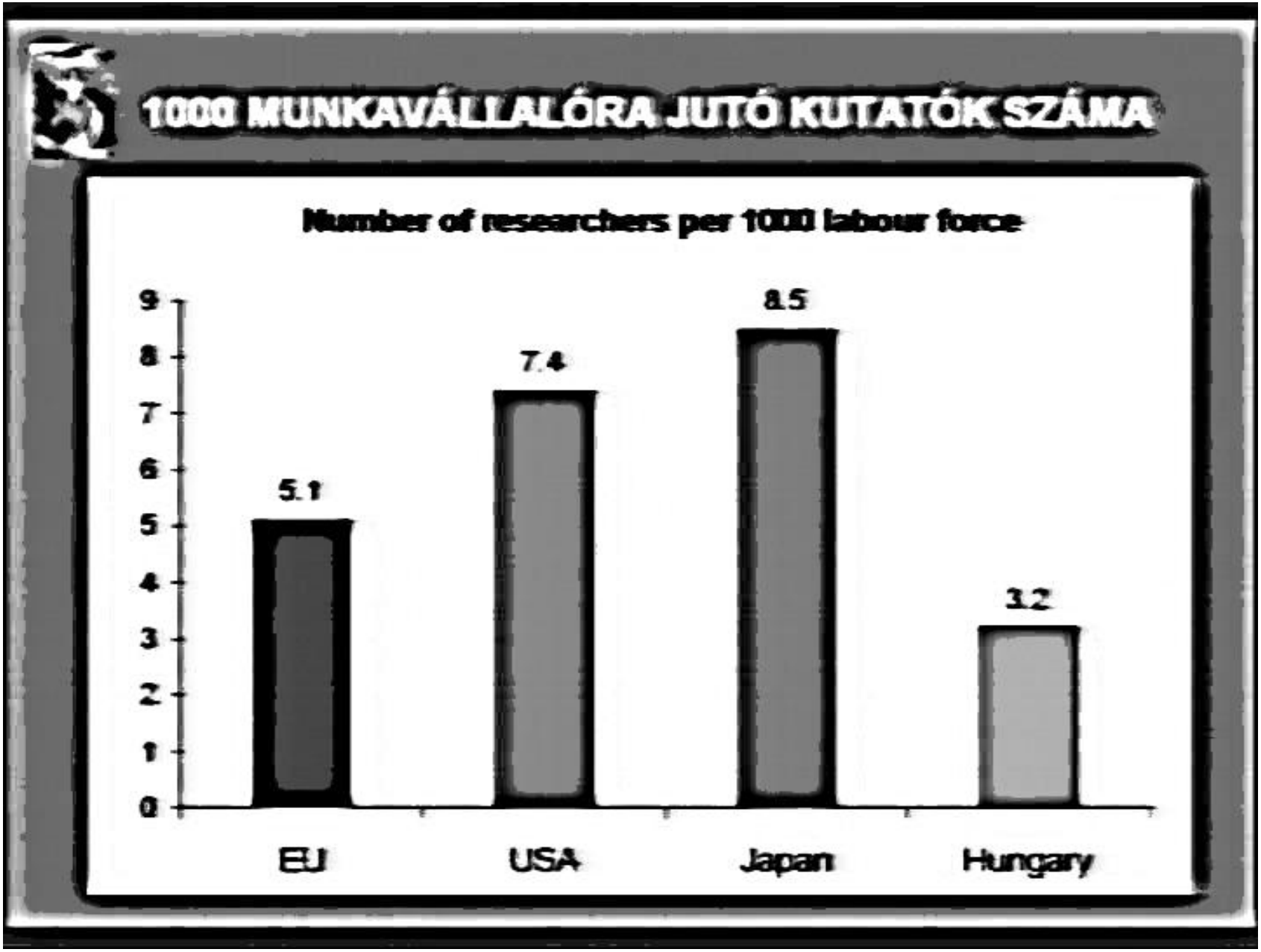
-unsharp 0x6+7%



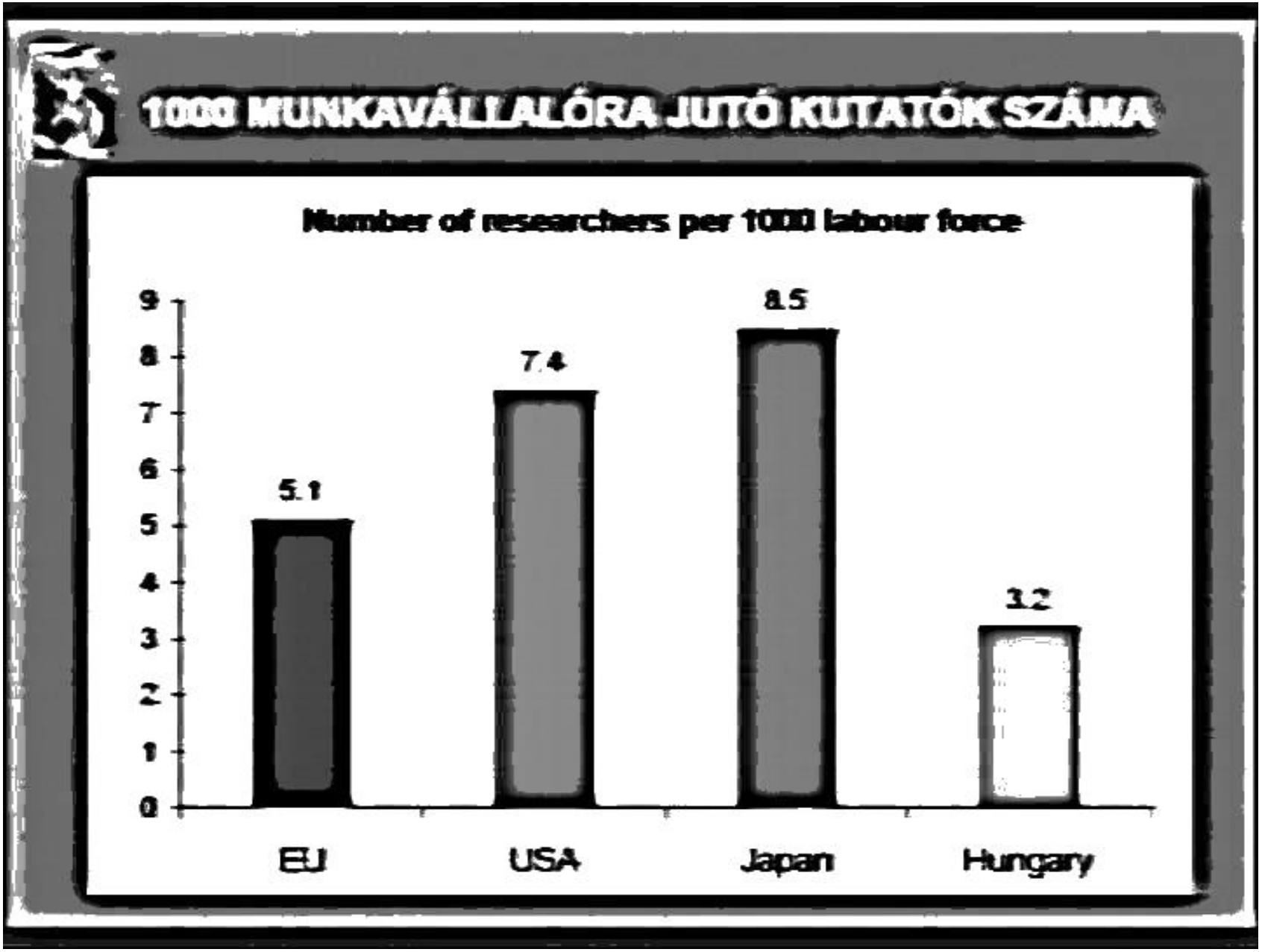
invertálás?


1000 MUNKAVÁLLALÓRA JUTÓ KUTATÓK SZÁMA
Number of researchers per 1000 labour force
9
8
7
6
5
4
3
2
1
0
5.1
7.4
8.5
3.2
EU
USA
Japan
Hungary

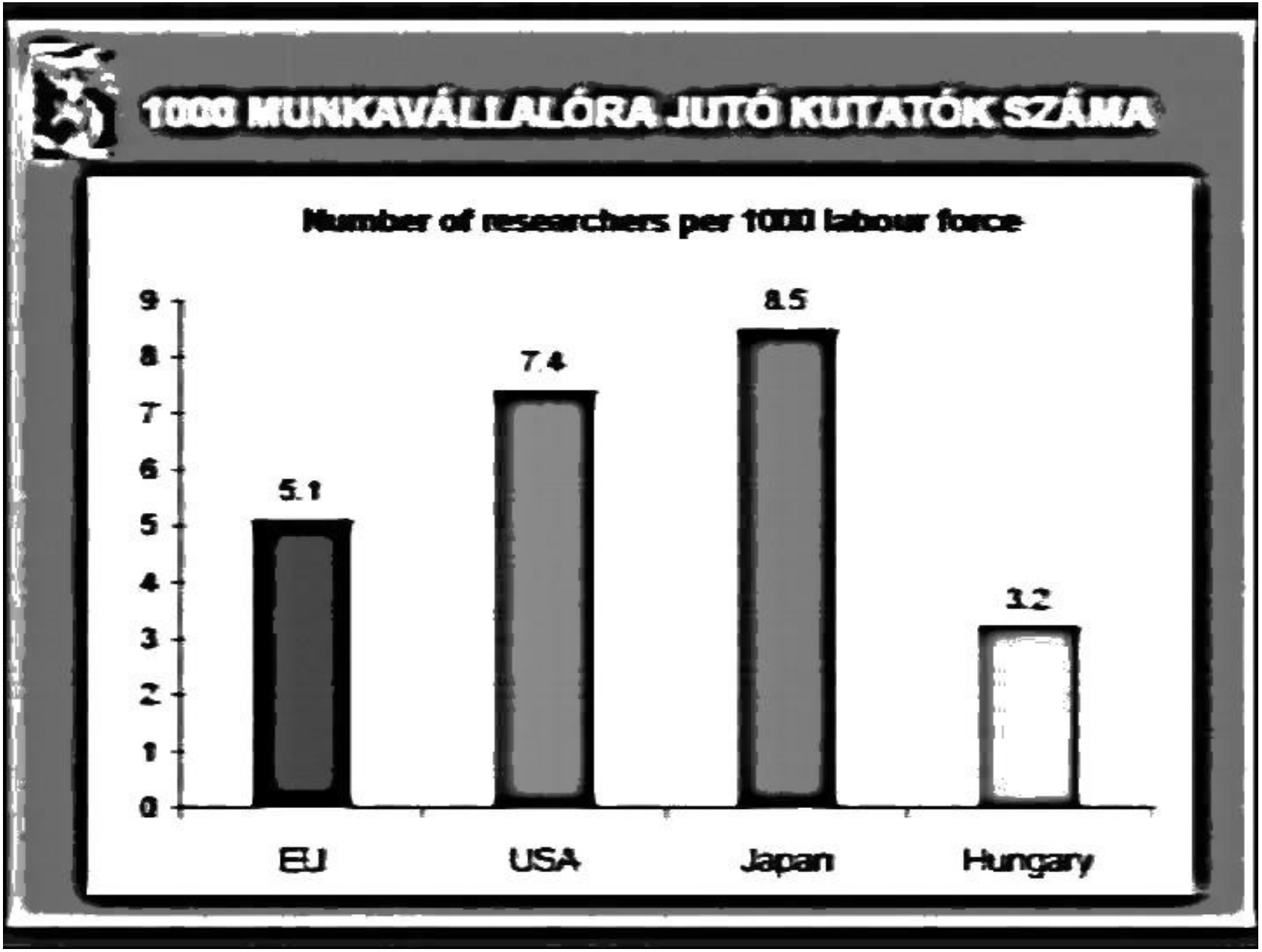
-contrast-stretch 0%x50%



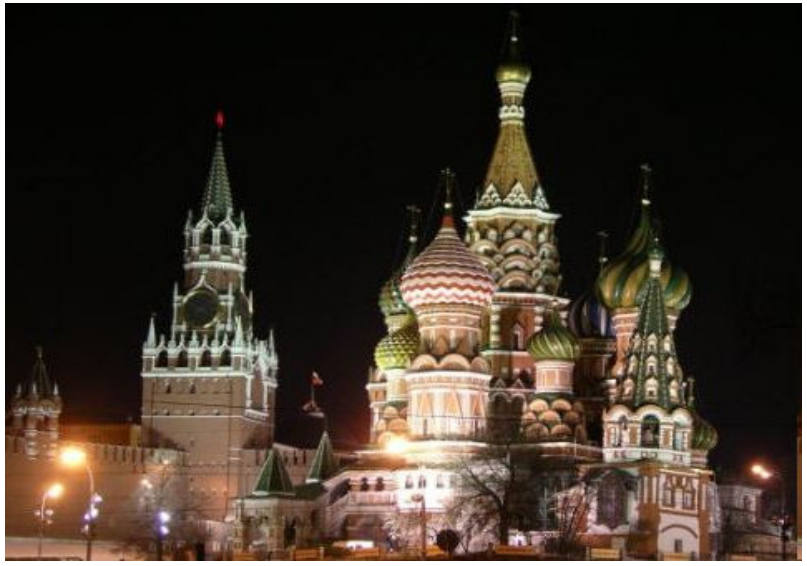
-white-threshold 68%

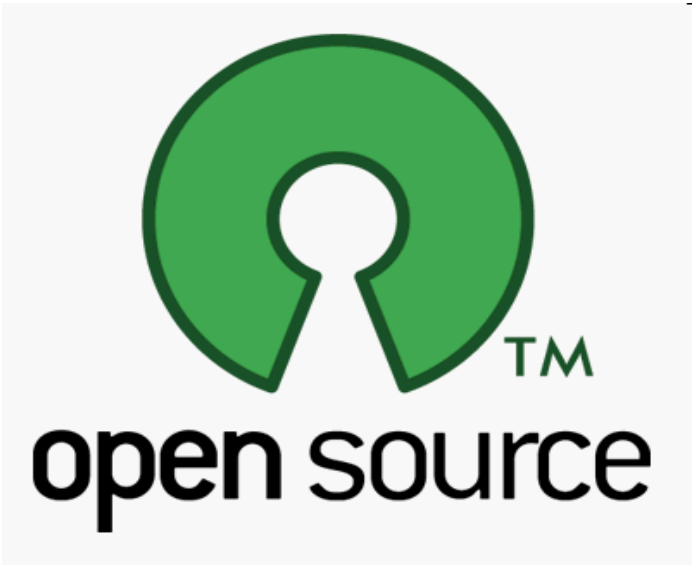


-despeckle


1000 MUNKAVÁLLALÓRA JUTÓ KUTATÓK SZÁMA
Number of researchers per 1000 labour force
9
8
7
6
5
4
3
2
1
0
5.1
7.4
8.5
3.2
EU
USA
Japan
Hungary

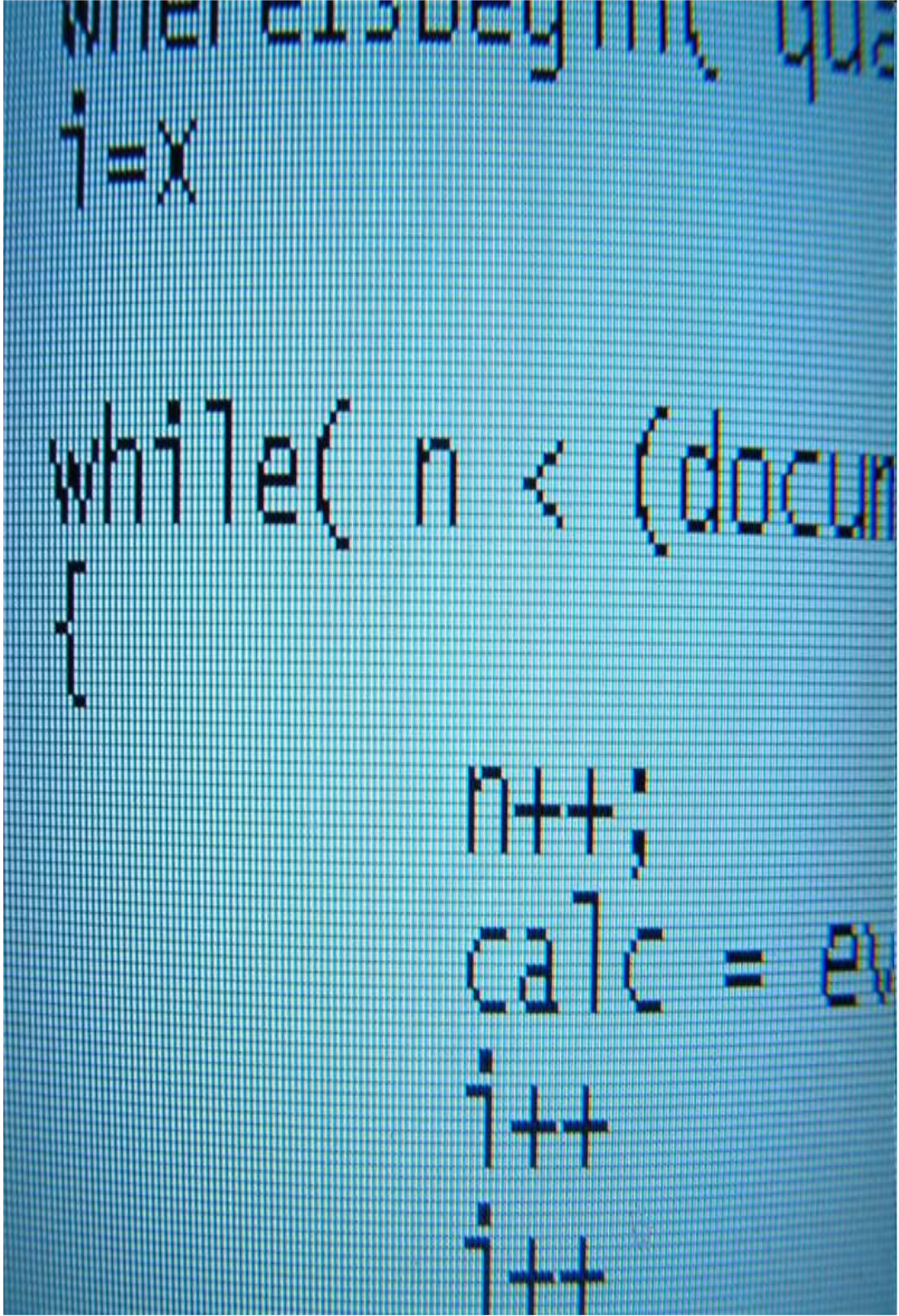
Cognitive OpenOCR (CuneiForm)





Fejlesztések

- előfeldolgozó és OCR konnektorok: cserél-hető komponensek
- elérhetőek a migráció és az alkalmazás számára is
- adatbázis módosítások



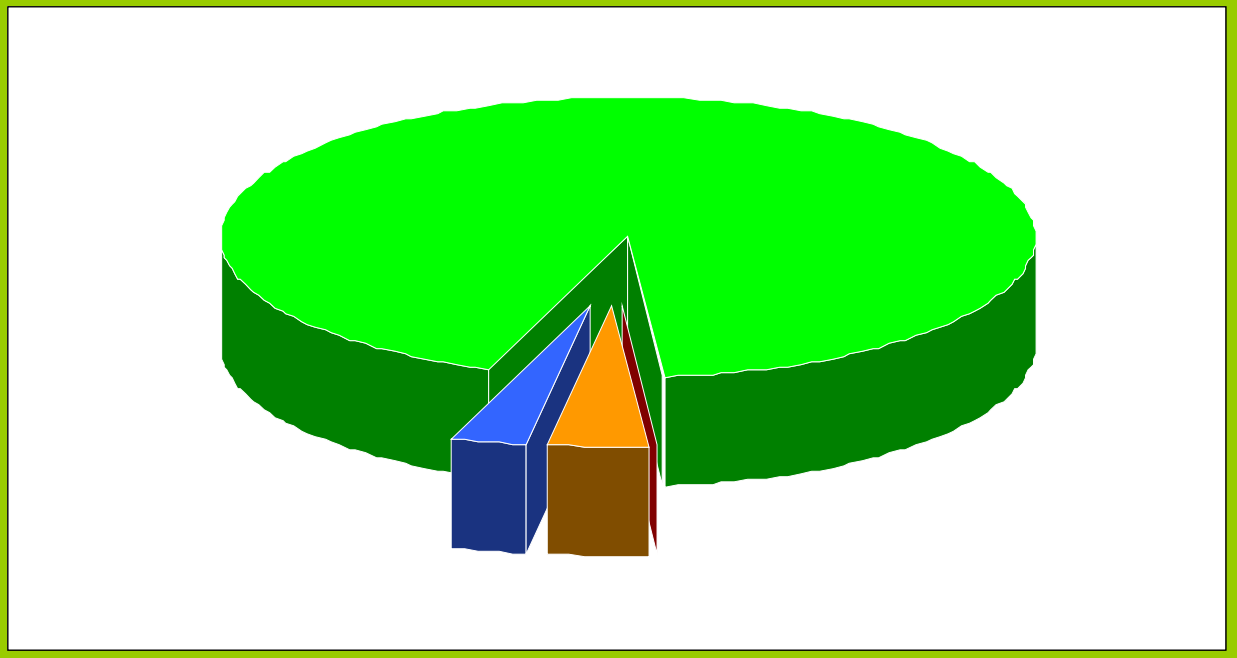
Rendszerbe illesztés

- feltöltött bitképek fogadása a szerkesztőfelületről
- előfeldolgozás
- OCR
- storage-re másolás
- részletes naplózás

Feldolgozás

- tesztkörnyezet kialakí-tása
- tesztfeldolgozások, javítások iterációval
- biztonsági mentés
- feldolgozás élesben
- kb. 1,9 GB
- 10-12mp/dia
- 6 nap

A feldolgozás eredménye



- **93%** - sikeres feldolgozás
- **3%** - korábban feldolgozva (PDF, PPT, ODP)
- **3,9%** - üres kimenet
- **<0,1%** - szoftverhiba

A kereső továbbfejlesztése

Felhők átvonulása és naplemente

2 órás felhőátvonulás Budapesten

Felvétel ideje: 2009. július 7. **Megtekintések:** 68 **Értékelés:** ☆☆☆☆☆

Találatok a fóliákon:


8mp


13mp


15mp


17mp


18mp


19mp


27mp

30mp

Tanulságok majdnem mindenkinek

- Célhoz az eszközt
- Ingyenes OCR még mindig körülményes
- Ár és érték sokszor együtt jár
- Kompromisszumkész-ség, kitartás!

Köszönöm a figyelmet!

Turcsányi Tamás
tt@niif.hu

http://videotorium.hu/