

# A véletlen, kiszámítás és információ viszonya

Benczúr András

ELTE IK

ELTE eScience RET

A Természet Világa alapításának 140.  
évfordulója alkalmából 2009. szeptember 18.



# Bevezető:

A kommunikáció, Shannon hírközlési modellje, az információelmélet kezdetei.

Az adatátvitel kérdése, adatátviteli hálózatok, műsorszórás.

A forrás bizonytalansága - más lett fontos: nem az elemi üzenetválasztás, hanem az összefüggő üzenet-adatvilág került előtérbe.

A múlt adatbázisának kezelése, a jövő adatbázisának tervezése összefonódik.



# Információelméleti háttér

- A múlt fuzzy, a jövő véletlen. A múlt van az adatbázisokban, a jövő lehetőségeinek alakítása van a számításokban: az eloszlást tudjuk befolyásolni, nem a teljes véletlent.
- Jim Gray: a mérések fuzzyk, a számítások közelítők – új tudományos módszerek, e-science
- Juris Hartmanis: a véletlen, algoritmusok, Kolmogorov komplexitás



Van már, amivel tudunk mit kezdeni, sok minden van, amire fejlesztik a technológiát.

Juris Hartmanis: (idézet Turing-díj előadásából, ACM Communications, 1994. október)

**„Meg vagyok győződve, hogy a számítástudomány már eddig is teremtett és továbbra is hatalmas lehetőségeket teremt fizikai és intellektuális világunk megértéséhez való hozzájárulásban. A kiszámítás paradigmája, egyre hatékonyabb univerzális számítási berendezésekkel ellátva, motiválja és teszi lehetővé a fizikai és intellektuális folyamatok feltárását és szimulálását, és egyben felbecsüli lehetőségeiket és korlátjaikat.”**



# Az információ mérőszámai

Kiindulás: mit mérnek?

Valaminek a leírását adjuk meg, **helyettesítünk** jelekből álló leírásokkal.

Lehetőleg rövid leírásokat keresünk.

A leírásból kívánt pontossággal visszanyerhető legyen, amit helyettesít.

Az információ mérőszámai leírások hosszára vonatkoznak, az adott feladat szempontjából optimális megoldás hosszát adják meg.



# Az információ mérőszámai

## Kolmogorov: Három megközelítés

1. **Valószínűségi:** Shannon-entrópia

$$H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i$$

2. **Algoritmikus:** Kolmogorov-entrópia

$C(x) = C_U(x) = \min \{ |p| \mid U(p) = x \}$ , és  $\infty$ , ha nincs ilyen  $p$ .

A definícióban  $U$  a rögzített referencia függvény,  
tipikusan az univerzális Turing - gépet választják.

3. **Kombinatorikus:** azonos hosszú kód a halmaz minden elemére



# Az feltételes Komogorov-entrópia

Definíció:

$C(x|y) \doteq C_U(x|y) = \min_{p \in U} H(p, y) \doteq x$  és  $\infty$ , ha nincs ilyen  $p$ .

A definícióban  $U$  a rögzített kétváltozós referencia függvény.

Prefix-változat:

Az  $U$  által használt kódok prefixmentes rendszert alkotnak.

A megfelelő entrópiák jelölése:

$K(x)$  és  $K(x|y)$ .

Az algoritmikus entrópiák konstans erejéig egyértelműek.



# Játék a kombinatorikus entrópiával

## $n$ kocka felcímkézése

Azonos hosszú bináris kódot kell ragasztani minden kockára:

$n \log_2 n$  bit összesen.

(Ennél kisebb összhossz prefixmentes kódokkal nem érhető el.)

Legyenek színes kockáink, és színenként kell azonos hosszú kódot ragasztani:

$\sum_{i=1}^k n_i \log_2 n_i$  bitre van szükség,

ahol  $k$  a színek száma, és

az  $i$ -edik színnel  $n_i$  számú kocka rendelkezik.

Mennyi bitet nyertünk:

$$n \sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n}{n_i} = nH\left(\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_k}{n}\right)$$





# Az egyenletesség szerepe

## A halmaz szerinti prefix entrópia

A kombinatorikus entrópia – az egyenletes eloszlás optimális kódja - miatt a halmaz szerinti feltételes prefix entrópiára teljesül a következő egyenlőtlenség:

Legyen  $S$  egy  $m$  elemű, természetes számokból álló halmaz. Ekkor:

$$\sum_{x \in S} K(x | S) \geq m \log_2 m$$

Az egyenletes hosszúságú kódnál átlagosan nem jobb a feltételes optimális kód.



$T_1, T_2, \dots$  - a standard felsorolása a prefix Turing gépeknek,  $T_i$  a  $\phi_i$  függvényt számolja ki.

**Definíció.** Adott  $y$  szerint  $x$  feltételes prefix Kolmogorov bonyolultsága:

$$K(x|y) = \min_{p,i} \{l(p) : T_i(y'p) = x, p \in \{0,1\}^*\}, i \in \mathbb{N}$$

$$K(x) = K(x|\varepsilon)$$

Az univerzális prefix TM :  $U(y'p) = T_i(y'p)$

Ekvivalens definíciót ad:  $K(x/y) = \min_q \{l(q) : U(y'q) = x\}$

Definiáljuk az  $A = \{x_1, x_2, \dots, x_m\}$  halmazra:

$K(A) = K(\langle A \rangle)$ , és  $K(x/A) = K(x/\langle A \rangle)$ .

- Honnan következtethetünk a jövő eloszlására? Hogyan befolyásolhatjuk a jövő eloszlását?
- A **múltra vonatkozó ismereteinket** kell felhasználni. Ez az ismeret most a **világháló adatbázisában** gyűlik.
- A múlt leírásának tömöríthetősége: a Kolmogorov entrópia világa
- A Shannon entrópia: a jövő leírásának tömörségére ad mérőszámot: a jövő lehetséges kimeneteleinek eloszlása ismeretében mi a megkülönböztető leírások hosszának várható értéke? Ennek alsó határa a Shannon entrópia. (**felkészülés a jövőre**)
- **Kétrészes kód**: törvényszerűség leírása és a maradó véletlen, tipikus adat (**jelentése a jövőre nézve**)



- A feltételes Kolmogorov bonyolultság alkalmazása, a két részes kód:
- a) Adott elemet tartalmazó egyszerűen leírható halmaz, abban tipikus elem
- b) Kevés elemszámú, egyszerűen leírható halmaz
- A halmaz lehet például egy valószínűség-eloszlásból származó tipikus minták halmaza (Bernoulli eloszlás)
- Példa:  $2n$  szögpontú,  $q, p$  paraméterű reguláris páros gráf

## Definíció: A Kolmogorov struktúra függvény:

$$h_x(\alpha) \stackrel{\text{def}}{=} \min \{ \log_2 |S| : x \in S, K(S) \leq \alpha \}$$

$$S^* : U \stackrel{\text{def}}{=} \langle S \rangle,$$

A tanú halmaz a  $h_x(\alpha)$  értékhez  $S_x(\alpha)$ .

$K(S_x(\alpha)) \leq \alpha$ , és  $\log_2 |S_x(\alpha)| = h_x(\alpha)$ .

$$K(x) \leq \alpha + h_x(\alpha) + c.$$

(i)  $S = \{x\}$ ,  $K(S) \leq K(x) + O(1)$ , tehát

$$h_x(K(x) + O(1)) = 0.$$

(ii)  $\alpha = 0 + O(1)$  esetében az  $\{y : l(y) = n\}$  halmazra

$$h_x(0 + O(1)) \leq n.$$

**Definíció:**  $x \in S$  esetén a halmazon belüli véletlenségi hiány

$$\delta(x|S) \stackrel{\text{def}}{=} \log_2 |S| - K(x|S^*), \text{ ha } x \in S \text{ és } \infty \text{ egyébként.}$$

**Definíció:** A minimális véletlenségi hiány függvény

$$\beta_x(\alpha) \stackrel{\text{def}}{=} \min_S \delta(x|S) : x \in S, K(x) \leq \alpha$$

az üres halmazon a minimum végtelen.

Az  $x$  kétrészes kódjához az MDL (Minimum Description Length) függvény a következő:

$$\lambda_x(\alpha) \stackrel{\text{def}}{=} \min_S (\log_2 |S| + K(x|S)) : x \in S, K(x) \leq \alpha$$

A három függvény között logaritmikus additív pontossággal (mind az értékekben, mind az argumentumokban) a következő egyenlőség teljesül:

$$\beta_x(\alpha) \stackrel{\text{def}}{=} h_x(\alpha) + \alpha - K(x) \stackrel{\text{def}}{=} \lambda_x(\alpha) - K(x)$$

# Az információ fogalma

Idegen szavak szótárából:

- Információ:
  1. felvilágosítás, tudósítás, tájékoztatás, hírközlés
  2. értesülés, adat
  3. tudósítási, tájékoztatási anyag, hír
- Informál: tájékoztat, felvilágosít, tudósít

# Saját megközelítésem

- Visszatérve az információ fogalmára, a három matematikai mérőszám behatárol egy fogalmat: az információ az, aminek a mennyiségét mérik. Más kontextusban használt információra ezek a mérőszámok felelőtlenül nem használhatók.
- *Szerintem az információ fogalma történetileg érthető,*
- Kezdet: Az **információ** az élettelen és élő határvonalának egyik oldalán: az **élőnél** jelenik meg. Az élettelenben az információ értelmezhetetlen.



# Saját megközelítés

- Az élő az információ reprezentálására nem csupán belső lehetőségeit használja, hanem az élettelen is felhasználja reprezentáció céljából. Hibás az a fogalomalkotás, ami ezt a reprezentációt az élettelen információjaként tünteti fel. Majd a kommunikációnál térünk vissza erre.
- Az információ a múltra vonatkozik, és a jövőre való felkészülésben hasznosul. Az eredeti latin szó egyik értelmezése sem vonatkozik a jövőre. A jövőről nincs tudósítás, a jövőre lehetnek előrejelzések, akár teljes biztosak is, de lehetnek csupán elképzelések is, akár minden esély nélkül.
- Valakinek a jövőre vonatkozó elképzelései is csak a múltból tudósítanak: egy múltban lezajlott gondolkodás eredményéről.

# Információ

Az élet, az élő egyed gyűjt információt, átad, örökít információt.

Miben, hol jelenik meg **először** az információ? A sejten belül, majd az élő szövet szerveződéseiben és mozgásaiban.

Miből keletkezik az információ: a fizikai-kémiai hatásokból, az érzékelésekből. Hogyan hat az információ: érzékelődik.

Ami érzékelődik, az mozgás (változás), és mozgást (változást) tud kiváltani. Az a mozgás megint érzékelődik is. És így tovább. **Van benne eleve rekurzió.**



# Információ

Az állatvilág egyik jellemzője a célirányos mechanikai mozgás, helyváltoztatás. Ehhez a mozgás vezérlésében, koordinálásában igen nagy teljesítményű információfeldolgozás, igen nagy "kiszámítási" teljesítmény szükséges. Az idegrendszer méretezése ehhez a csúcsteljesítményhez igazodik. Amikor nincs mozgás, ez a kapacitás kihasználatlan, előfordulhat, hogy más információfeldolgozást is végezhet. Akaratlagossá is válhat ez a tevékenység, bizonyos mértékben irányíthatóvá, ami a gondolkodáshoz vezet.

Ennyi dióhéjban a biológiai individuumok belső „informatikája”.



# Információ

A következő informatikai szint az élővilágban az egyedek közötti kommunikáció. Ennek kiinduló elve az **idegrendszer függetlensége**. Ez alatt azt értem, hogy egyik élőlény a másik belső információreprezentációit és folyamatait közvetlenül nem ismerheti. Nincs az idegrendszereknek közvetlen együttműködése. Az egymásra hatás csak közbeiktatott közeg változásain keresztül történhet. Ez a kommunikáció.

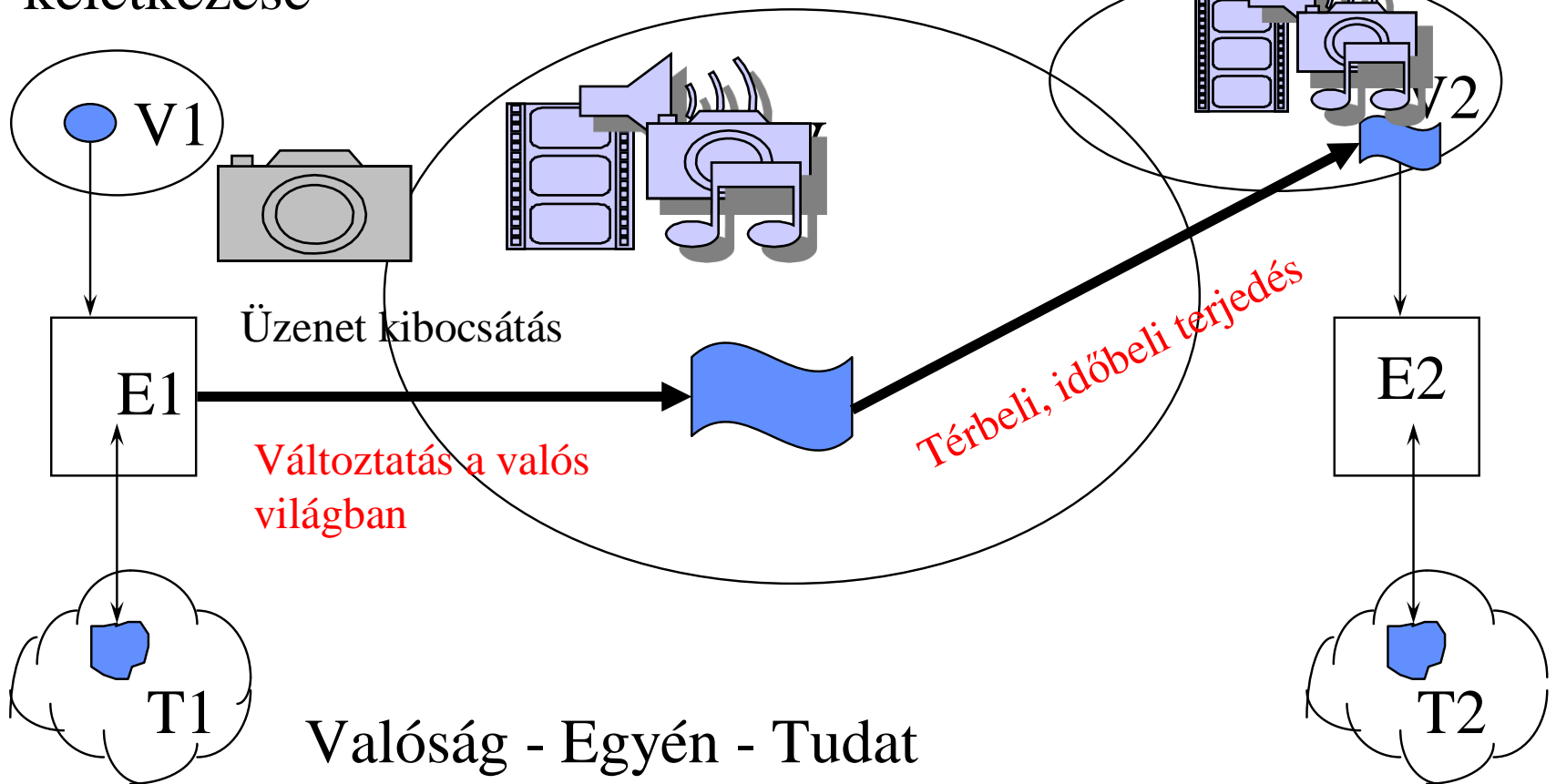
Eddigi írásaim innen indultak: a kommunikációból.



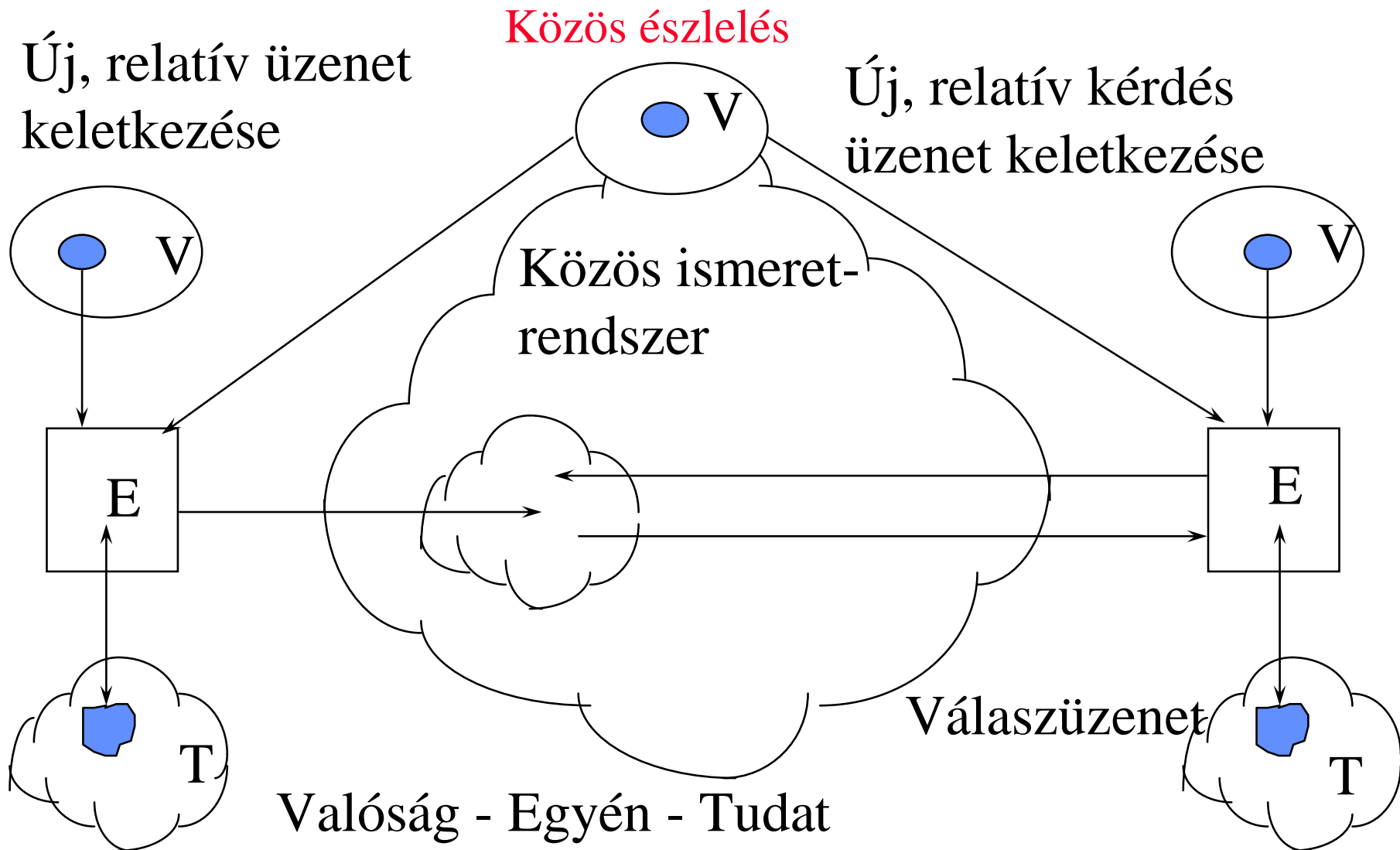
Új hangsúly : instrumentális elemi észlelés, élménymegosztás,  
vizuális üzenetek, tetszőleges multimédia forrás használható

Új üzenet  
keletkezése

Üzenet észlelése

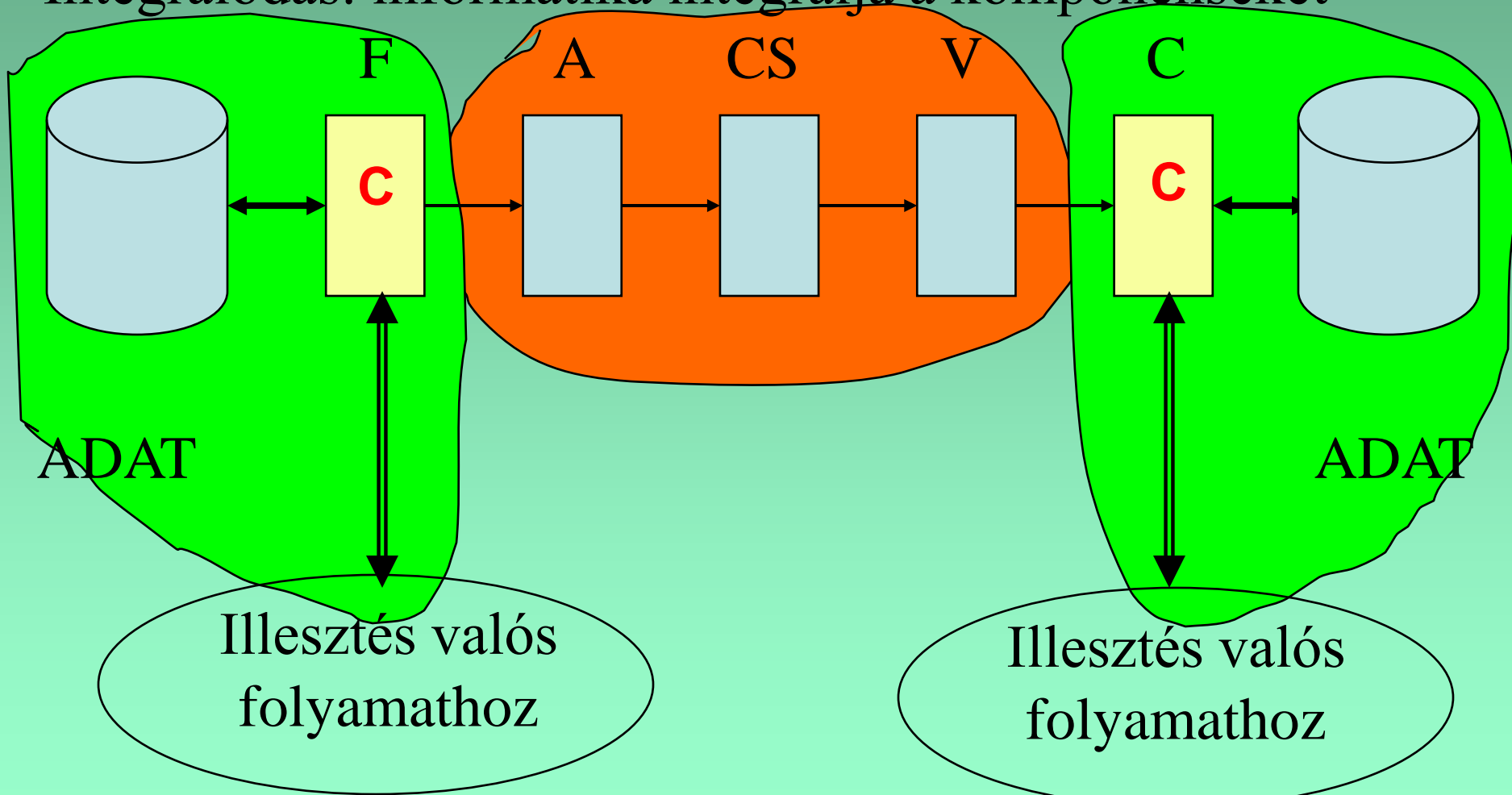


# Információs rendszer



Konvergencia: informatika-távközlés-média

Integrálódás: informatika integrálja a komponenseket



Multimédia, mérés és automatizálás, folyamatirányítás, térinformatika, bioinformatika, e-világ, CAD, CAM, ...

# A véletlen és kiszámítás

információra nem lenne szükség véletlen nélkül, az információ nem lenne használható kiszámítás nélkül. Nem lenne semmi szerepe. A véletlen matematikai modelljét a jövő leírásának eszközeként használhatjuk, mint a jövő bizonytalanságának, kiszámíthatatlanságának leírási módszerét.

A múlt homályos leírása és a múlt véletlen gyökerei, valamint a jövő bizonytalanságát modellező valószínűség-eloszlásra való következtetéseink hogyan hozhatók össze?

A jövőre csak a múltból következtethetünk.

Beck Mihály a „Parajelenségek és paratudományok” c. könyv 67-oldalán: „ A különböző folyamatok időbeli lejátszódásának leírása valójában csak akkor lehetséges, ha a jövő nem más, mint *megismételt múlt.*”





Mindez erőteljesen összefügg a kiszámíthatóság világával, s ezért ide tartozik a Turing-díjas Juris Hartmanis (az idő- és tárkonyoltság elméletének egyik megalapozója) Turing-előadásából:

**„A kiszámítás paradigmája lehetővé tette a véges és végtelen sorozatok Kolmogorov bonyolultságának segítségével a véletlen fogalmának tisztázását.** Megint, a kiszámítási modell univerzalitása lényeges szerepet játszik ezen fogalmak érvényességének bizonyításában.

Számítás-bonyolultsági megfontolások finomították a véletlenség viszonylagos fogalmát a tekintetbe vett alkalmazáshoz viszonyítva. Bizonyítást nyert, hogy mi véletlen (elfogadható vagy átmegy a teszten), az attól függ, mekkora a számítási kapacitása az adott alkalmazásnak. Továbbra is igen mély nyitott problémák vannak ezen a területen a fizikai folyamatok és a kiszámítás viszonyáról. A Kolmogorov-véletlen sorozat (nem tömöríthető egyik kezdőszelete sem) nem kiszámítható, mégpedig igen erős értelemben: nincs olyan Turing gép, amely ki tudna írni egy véletlen sorozatot saját méreténél (program hossza, vagyis a leírása) nagyobb hosszúságban. Vajon a megfelelő törvény (tétel) igaz-e minden fizikai rendszerre? **Tud-e egy kis fizikai rendszer tetszőleges hosszúságú Kolmogorov-véletlen sorozatot előállítani, vagy még pontosabban, képes-e egy véges fizikai rendszer (pontosan meghatározva a szükséges energia bemenettel, de véletlenszerűség hozzávétele nélkül) korlátlan Komogorov-véletlen sorozatot előállítani? Ha igen, akkor valójában a fizikai folyamatokat nem lehet teljes mértékben szimulációval jellemezni.”**



# Rényi Alfréd egy kérdése

**„Lehet egy vizsga nehézségét azzal jellemezni, hogy hány bit-et kell a hallgatóknak tudni? Enciklopédikus jellegű tárgyakban ez nem is teljesen abszurdum, a matematikában, persze, ennek nincs értelme, hiszen a dolgok egymásból következnek, aki az alapokat tudja, elvben mindent tud, illetve tudhatna. Egy matematikai elmélet összes eredménye tulajdonképpen csírájában benne van az axiómákban – vagy mégsem? Erről egyszer még gondolkodni fogok.”**

(Rényi „Az információ matematikai fogalmáról” (Egy egyetemi hallgató naplója) Ars Mathematica, Rényi Alfréd összegyűjtött írásai, TYPOTEX, 2005.)



# A válasz az algoritmikus információelméletben van.

Egy matematikai elmélet összes (bizonyítható) eredménye felsorolható, csak győzzük kivárni, mikor érünk el az éppen kért eredményig. A feltételes Kolmogorov-entrópia mutatja, hogy adott bitnyi ismeret matematikai tétel esetén kevesebb kiegészítéssel vezet el egy kérdés válaszáig, mint történelmi tétel esetén, másként mondva, kevesebb újonnan megtanult bit feldolgozása után érünk el a tételhez. Azonban a feldolgozás sebessége sem elhanyagolható. Sőt, a felhasználható tárméret sem. Fejben más a határ, mint papír-ceruzával.



# Paradoxon

Paradox módon az algoritmusos információelmélet szerint a matematika egyszerűbb, mint a történelem. Intuitív magyarázata ennek az állításnak az, hogy ugyanannyi mennyiségű írásos válasz a vizsgán (matematikából a definíciók, tételek pontos megfogalmazását és a tételek bizonyításának leírását értve válasz alatt) kevesebb memorizálandót jelent matematikából, mint történelemből. Ugyanígy magyarázható, miért egyszerűbb a vers, mint a próza megtanulása szó pontossággal. A matematika kevésbé véletlenszerű, mint a történelem. (Még irritálóbb állítás, hogy az élő anyag egyszerűbb, mint az élettelen. Ennek kifejtése a genetika-genomika világába vezet már. A genetikai kódnak a bioszféra gépezetébe helyezésével történő kiszámításként elképzelve az egyed kifejlődését.)



# Turing-gépekkel illusztrálva

Próbáljuk meg Turing-gépekkel illusztrálni a vizsgát.

(A Turing gép. A véges leírás jelentése, viszonyítás a generatív jellegű parciális rekurzív függvények építkezéséhez.)

A megtanult bitmennyiség legyen  $M$ , mint Magolás. Ez a Turing-gép egyik szalagjára kerül. A Turing-gépnek legyen a magolt résztől elkülönülten kialakult programszalagja is.

(Ez modellezi az előzetes tanulás során kialakult gondolkodást.) Ezen kis Turing-gépek programjai

(mozgástáblái) vannak.



# enciklopédikus tudás

Mi az **enciklopédikus tudás jellemzője**: Az M eléréséhez elsősorban indexelés kell, és elért szekvenciák visszaadása. (Az asszociatív emberi visszakeresés is sajátos indexelés. Ebben is nagy a gyakorlás szerepe, és a korábban, más céllal megtanultak szerepe.)



# matematikai tudás

Az axiómarendszerre fogalmak és erős szerkezetek épülnek. A szerkezetek egy része definíció és tétel jellegű, vagyis az axiómákból következő igaz állítások megfogalmazását könnyítő definíciók és utána tételek megfogalmazása. A tételek igazolása algoritmusokkal történik. Turing-gépnél az algoritmusok egy része bitekkel adott a programszalagon, egy része az M-ben új adatként van. A kettő erős együttműködése tesz lehetővé kevés új bit megtanulásával nagyobb visszaadható tudást. A tételek generatív megfogalmazhatósága (szabályok alapján állíthatjuk elő tételek sorát) rávezet a tételek megfogalmazására pontatlan tudás esetén is, ugyanígy a bizonyítások is generatív-felsorolható módon rekonstruálhatók.





Azonban a rekonstruáláshoz igen gyors számítási teljesítményre van szükség. Gondolkodni és érteni kell! A gondolkodást gyakorolni kell. **(Konfuciusztól idézet emlékezetből: Tanulni csak gondolkodva érdemes. Gondolkodni tanulás nélkül veszélyes.)** A tételek generatív előállítás az elméletek formális nyelvként való megadhatóságára épül. A tételekből egy elég sűrű részhalmaz tudása alapján könnyű kitölteni sok további tételt. Ehhez viszont jól kell ismerni a nyelvet. A tételek és bizonyítások szövedéke újabb tételekhez vezet. A matematikai vizsgán a „ritka” tételtudás esetén gyors generatív, „kiszámító” gondolkodás is elég lehet, a sűrűbb tételhalmaz tudása viszont gyengébb gondolkodási képesség mellett is elegendő lehet.



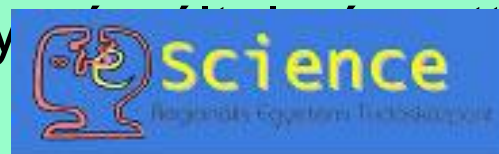


. A matematika múltja igen sok lehetséges ág bejárását jelenti. Ez is a megismert múlt része, a megismert gondolatok múltja. Ebből építkezünk tovább. A matematika maga – lehetséges axiómarendszerei, tételei – nem változik. Az változik, hogy ebből emberi tudatok mit jártak be. A matematikusok teremtett világai ebben a legnagyobb kirándulások. A teremtés új és új terminális és grammatikai elemeket hoz be. Új nyelveket kell tanulni. Egyre többet lehet bennük megfogalmazni. **A megfogalmazottak (megismert tételek) mennyisége eléri-e az emberöltő által továbbdolgozható volument? Osztott, párhuzamos és ellenőrizhető matematikai gondolkodás jön-e majd létre?**



# Informatikai tudásanyag

Az informatika maga, és ezért ismeretanyaga változik, és igen gyorsan. (Szemben a természettudományokkal, ahol nem a természet változik, hanem ismereteink és észlelési lehetőségeink bővülnek.) A változást emberi tudatok bejárása eredményezi. Teremtett világ. Vannak benne matematikai jellegű részvilágok, és tele van heurisztikákkal is. Az eredmény: programok, technológiák, materializált gépek. Utána azonban gépek folytatják a lehetséges „tételek” bejárást. Ezekről a bejárásokról megint emberi tudatnak kell tételeket fogalmazni. Mire a tétel kész, már lehet, hogy más bejárások (programok, technológiák) leváltották a tétel tárgyát képező bejárásokat. (A bejárás itt tulajdonképpen kiszámításokat jelent, vagy emberi tudat által vég



Korábban voltak a **könyvek**, oda ömlött be minden tudásunk (helyesek és hibásak, jó szándékúak és rossz szándékúak.)

Vörösmarty: „Gondolatok a könyvtárban”

Az USA Kongresszusi Könyvtár: **28 millió könyv**. Teljes digitalizálása: 10-100 Mbyte/kötet:összesen **280-2800 TeraByte**.

Most a világháló adatbázisába ömlik minden. Becsült mérete **Zetabyte** tartományban van. Másfél évenként kétszereződik , gyorsul.

**Exabyte**: az 1999-ben keletkezett információ (adattömeg) fele.

IDC tanulmány szerint közel fél Zetabyte, pontosabban **3 892 179 868 480 350 000 000**

bitnyi információ keletkezett 2008-ban, 2009-ben ezerszer annyi várható, mint 1999-ben.



# Programkódok

Mennyi az elkészült programok bitmennyisége? Hogyan viszonylik ez a világháló  $10^{21}$  (Zeta-bájt) méret közelében lévő adatmennyiségéhez?

Megbecsülhető, hány bájt keletkezhet 100 millió programozó napi 16 órás munkájával másodpercenként egy leütés (fél bájt) írási sebességgel évente:

Az eredmény: 1.051.200.000.000.000 bájt.

Egyszerűsítve, ez 1 Exabájt. Ez a szinte irreálisan magas felső becslés évente ekkora növekedési korlátot mutat. (Ha 100 millió digitális kamerát működtetnénk másodpercenként egy felvétellel, az évente kétmilliószer ekkora adattömeget jelentene, 2000 Zetabyte lenne az eredmény.)



# Tudományos adatok

Milyen arányt képviselnek?

A legnagyobbak Petabyte tartományban.

CERN: évente 10Pbyte

A világháló nagy része azonban a közvetlen emberi érzékelésnek megfelelő észlelések rögzítéséből áll.



# Méretarányok

A háttértároló 100-szor olcsóbb, mint a gyors memória, igen nagy tárolókapacitás jött létre. (60% PC-ken van)

**Amdahl törvények** a kiegyensúlyozott rendszerről: a másodpercenkénti lebegőpontos műveletek száma, az operatív memória mérete byte-ban, a **másodpercenkénti I/O mennyiség byte-ban**, valamint a háttértároló mérete úgy aránylik egymáshoz, mint **1:1:0,1:100**.

A petaművelet/sec tartományban 1 GByte/sec adatelérési sebességű merevlemezekből 100000 kell, 1 TeraByte kapacitású tárolókból 100000.

Visszafele: 1 Zetabyte adatbázishoz 10 Exaflops processzor telejsítmény, 10 Exabyte memória, 1 Exabyte beolvasási sávszélesség kell.



# Cyber-infrastruktúra

Z. Karvalics László

A cyber-infrastruktúra mint aktuális kihívás és mint tudományszociológiai probléma I. Magyar Tudomány 2007.

A cyberkörnyezet egyaránt támogatja a tudománygyár működésének átfogó újratervezését (*re-engineering*) és a kutatási folyamatok jobb programozását - evvel **a tudomány új korszaka** születik meg (*next generation science*), amit bátran nevezhetünk **adat-intenzív tudománynak** (*data intensive science*)

(Jim Grey, Alex Szalay et.al. 2005).



# Jeltömeg és kontrollválság

Z. Karvalics László

Az adatsilóktól a tudomány kontrollforradalmáig.  
Magyar Tudomány 2008.

*„A tudósok lényegesen gyorsabban hozzák létre az új adatokat, mint ahogy azokat elemezni tudnák. Az eredmény leginkább az optikai csalódásra hasonlít”.*

*(Hugh Kieffertől származó idézet)*





# Mi az e-science?

Az ELTE eScience RET pályázatból

Új feltörekvő technológia, melynek révén nagyléptékű, komplex tudományos tevékenység fejthető ki a modern információs technológia felhasználásával. Legfőbb jellemzője a rendkívül sok, gyakran különböző helyekről elérhető adaton operáló kiértékelő munka, melynek eredményes véghezvitelére az adatok automatikus gyűjtése, optimális adatbázisba rendezésére, rendkívül nagy számítástechnikai kapacitást igénylő feldolgozására, és a lényegret megragadó vizualizációjára van szükség.



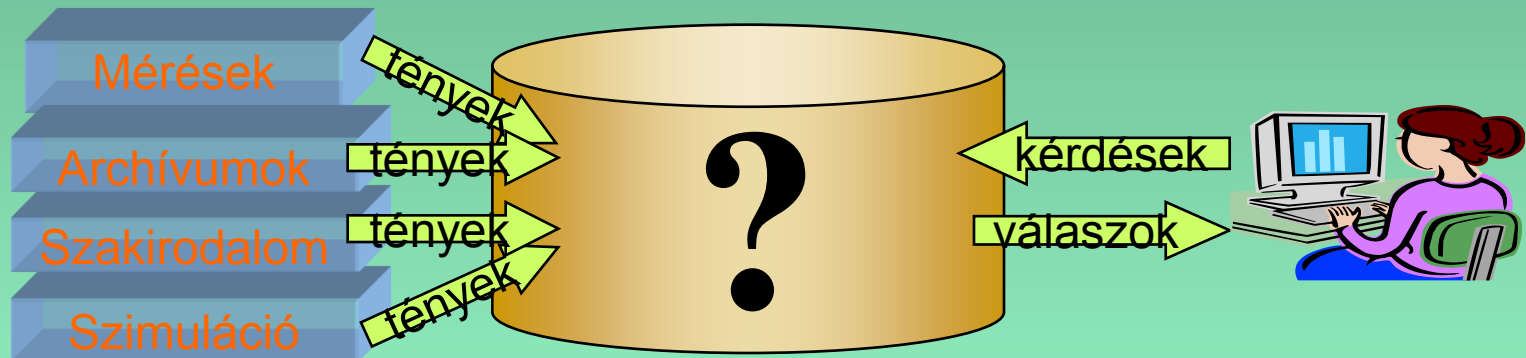
# CyberinfrastruCture Vision for 21st Century DisCoVery

National Science Foundation  
Cyberinfrastructure Council  
March 2007





# Az átfogó kép a tudományban



- Adatfeltárás
- Petabyte kezelése
- Közös séma
- Hogyan szervezzük?
- Hogyan szervezzük át?
- Hogyan működünk együtt?
- Adatlekérdezés és vizualizáció
- Hatékonyság

# <http://SkyServer.sdss.org>

- Solves the 20 queries
- Has 150 hours of online instruction
  - Translated to Hungarian



- Professional astronomers use it as the SDSS Science Catalog Analysis Service.
- Clone operating in Hungary.





# Experiment Budgets $\frac{1}{4}$ ... $\frac{1}{2}$ Software

## Software for

- Instrument scheduling
- Instrument control
- Data gathering
- Data reduction
- Database
- Analysis
- Visualization

## Millions of lines of code

Repeated for experiment  
after experiment

Not much sharing or learning

**Let's work to change this**

## Identify generic tools

- Workflow schedulers
- Databases and libraries
- Analysis packages
- Visualizers
- ...

Simulation (computational science) are  $> \frac{1}{2}$  software

# Észlelés – megismerés – információ

Fizikai jelenségek észlelése: egy másik, már ismertebb jelenségre való hatáson keresztül történik. A kezdeti észlelés, mint felismerhető jelenség, az emberi érzékszervekre, érzékekésre alapul, a telereceptoraink, azaz a látó és halló receptoraink informálnak bennünket..



# helyettesítés

Minden észlelésből helyettesítő jelek maradnak. Tudatunkban is. Utána már csak ezekkel tudunk manipulálni.

Ez viszont csak algoritmusokkal történik – absztrakt értelemben. Az emberi gondolkodás, következtetés, stb, minden eddigi lehetősége a kiszámítási világon belül van. Ami ezen túl van, az nem ellenőrizhető. (Bár a kiszámítás, gondolkodás se mindig



ellenőr





# megismerés

Miben áll a természettudományos megismerés legfontosabb lehetősége: Addig kell provokálnia természetet, míg olyan új múltat nem eredményez, amelyet eddig még nem észleltünk. A jövőre csak a múltból következtethetünk.

Ezt fogalmazza meg másként Beck Mihály a „Parajelenségek és paratudományok” c. könyv 67- oldalán:  
„ A különböző folyamatok időbeli lejátszódásának leírása valójában csak akkor lehetséges, ha a jövő nem más, mint *megismételt múlt.*”



# Zárszó

## Juris Hartmanis: Zárszó:

„Hiszek abban, hogy a számítógéptudomány rendelkezik olyan potenciális erővel, amely által mélyebben be tudunk tekinteni a kiszámítás paradigmájába, valamint saját intellektuális folyamatainkba, kvantitatív megértésüket kaphatjuk, és így, esetleg talán, egy lehetőséget nyerünk a tudható határának átlépésében.”



# Köszönöm a figyelmet!

