



Meaning Based Computing

IMDB @ Adatb haladóknak

Wagner Kornél
kornel.wagner@hp.com



Adatbázisok haladóknak 2012.

2012. október 2.



Miről lesz szó?

Információ kezelés

Meaning Based Computing

Autonomy

Alkalmazási lehetőségek

Kérdések, válaszok

Name: Smith

Address: 46 Acacia avenue

Age: 32

Code: 00403034

Reference FFRRF55434
=====



A világ nem sorokból és oszlopokból áll

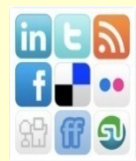




Információ

	Strukturált	Nem strukturált
Mennyiség	15%	85%
Növekedés	22%	62%

- A nem strukturált információt a számítógépek nem tudják értelmezni
- Például:



Közösségi
média



Video



Hang



Email



Szöveg



Dokumentum



Kép



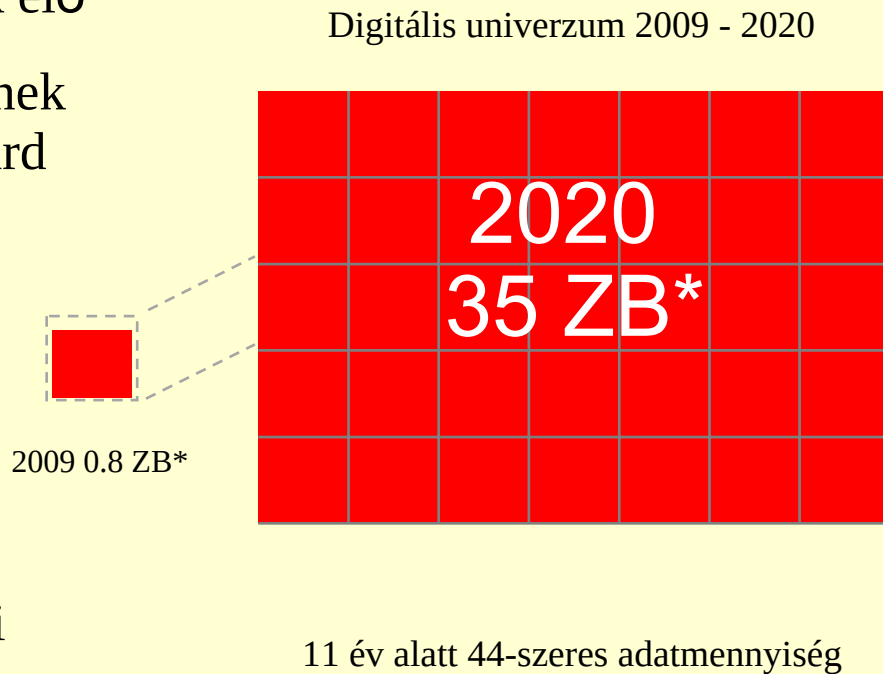
Információs trendek

- Közösségi média
 - A cégek 58%-a használja a közösségi médiát, 21% pedig tervezi a használatát. Viszont nincsenek felkészülve a hatékony kezelésére, csak 12% elégedett a tevékenységével
- Big data
 - A méret nem akadály. Az igazi probléma az analízis hiánya.
- Cloud
 - A cégek 77%-a tervezi a cloud infrastruktúra használatát. 9 éven belül a tranzakciók fele már Cloud infrastruktúrán fog történni.
- Mobil
 - A Fortune 500 cégek 88%-a használ mobil alkalmazásokat.



A digitális információ mennyisége

- Az információ 70%-át emberek állítják elő
- Minden percben 35 órányi videót töltenek fel a YouTube-ra. Naponta kb. 2 milliárd videót néznek meg.
- 2010-ben 5 milliárd mobiltelefont használtak
- Minden hónapban kb. 30 milliárd bejegyzést hoznak létre a Facebook-on
- 2011 áprilisában az USA Kongresszusi Könyvtára 235 TB adatot tartalmazott
- A digitális adatmennyiség becsült éves növekedése kb. 40%, ezzel szemben a globális IT költség évente kb. 5%-al növekszik.



*Zettabyte = 1 millárd TB



Integrated MBC

Clustering, Scene Detection, Identification & Sentiment

Categorize, Alert, Profile & Hyperlink

Audio & Video

Conceptual Search

Legal Search

Secure Search

Federated Search

Structured & Parametric

Page Rank

Keyword Search+

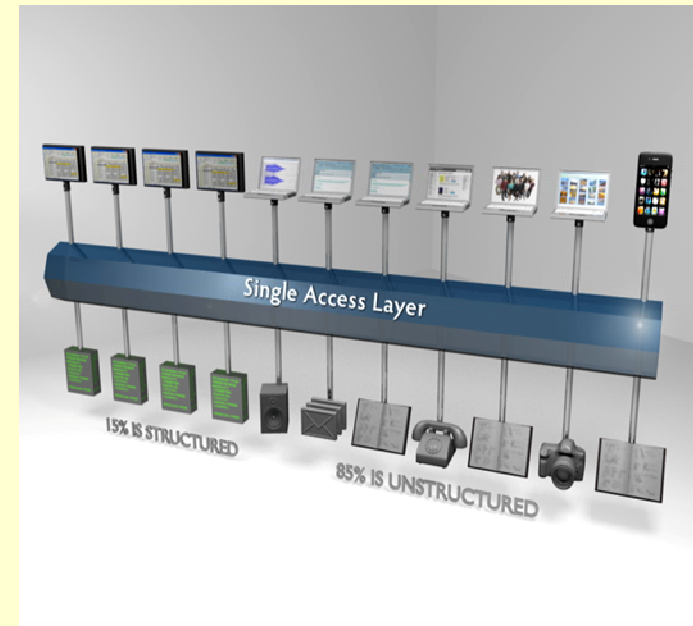
Keyword Search





Meaning Based Computing

- A strukturált és nem strukturált információ együttes kezelése
- Elosztott, különböző helyeken elérhető információ kezelése
- Az összes információ jelentésének és összefüggéseinek értelmezése
- Valós idejű analízis
- Az információ publikálása az alkalmazások számára



Autonomy



Real-Time Analytics

Agile -Environment

In-Memory

Unlimited Scale

Agents

Community

Interest Profiling

Expertise Profiling

Retrieval
Concept
Keyword
Parametric
Federation

Summarization

Hyperlinking

Channels

Clustering

Categorization

Classification



Matematikai alapok

- Bayes-tétel
 - “A” valamiféle hipotézis, “B” egy megfigyelhető esemény. A tétel azt adja meg, hogyan erősíti vagy gyengíti az esemény megfigyelése a hipotézis helyességébe vetett hitünket.
 - Pl.: 100 fej, Monty Hall-paradoxon
- Shannon-tétel: $H = -\sum p_i \log_2(p_i)$
 - A jel információmennyisége függ a jel előfordulási gyakoriságától.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$



Strukturált adatok kezelése

- Hagyományos kezelés
 - Adattárház, BI
- SQL helyett keresés
 - Mezők kialakítása
 - SQL szerű keresési feltételek
 - SQL-en felüli eszközök: pl. súlyozás
- Valós idejű analízis
 - Egyedi lekérdezésekre jó
 - OLAP kockát nehezen helyettesíti



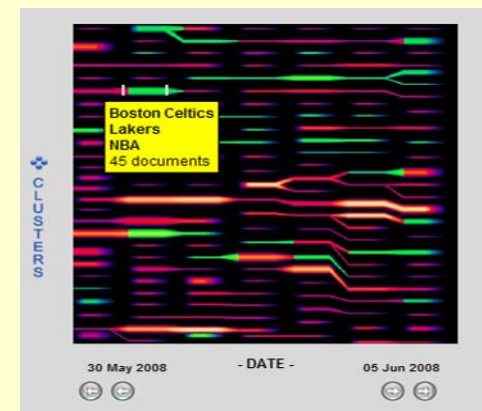
Nem strukturált adatok értelmezése

- Keresés
- Kivonat készítés
 - A kulcsfogalmak és szövegkörnyezetük kiemelése a szövegből.
- Hiperlinkelés
 - Indexelési időben dokumentumok összekapcsolása



Információk csoportosítása

- Csoportosítás (klaszterezés)
 - Minta felismerés: A dokumentumban előforduló szavak és gyakoriságuk elemzése
 - Fogalmak kialakítása
 - X% a valószínűsége, hogy az adott fogalommal foglalkozik a dokumentum
- Kategorizálás
 - Kategóriák létrehozása és a dokumentumok kategóriákhoz rendelése
- Csatornák
 - Témakör dokumentumainak összegyűjtése





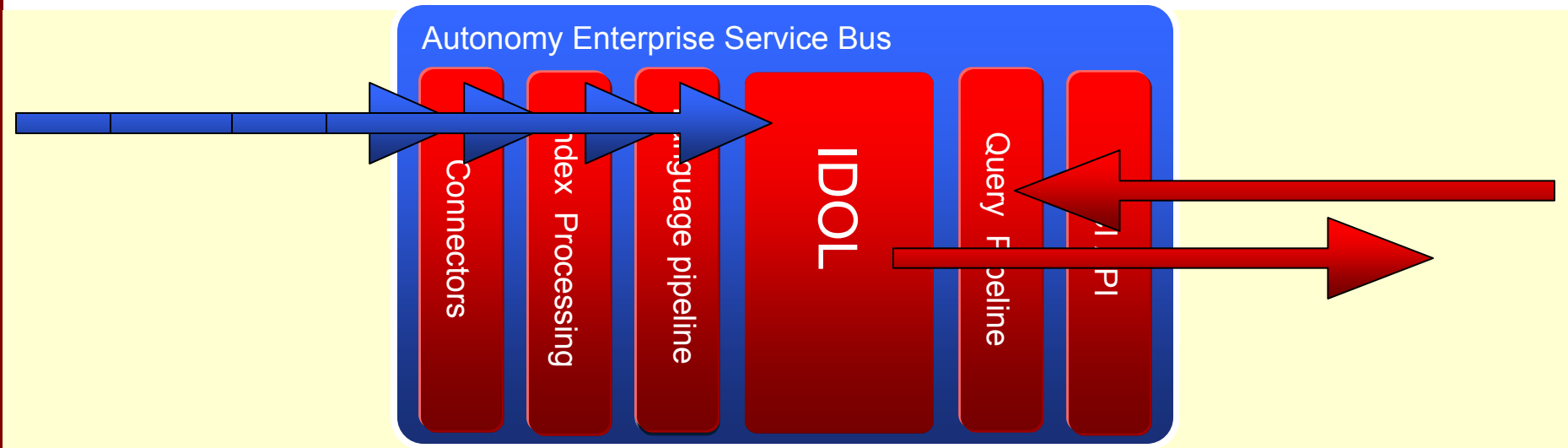
Felhasználók kiszolgálása

- Agent-ek definiálása
 - Érdeklődési körnek megfelelő dokumentumok összegyűjtése
- Alert-ek
 - Új, változott dokumentumokról értesítés
- Profile-ok kialakítása
 - A felhasználó érdeklődési körének meghatározása



IMDB @ Adatao haladóknak

Autonomy architektúra



500+ Functions

- ATG Dynamo
- BEA
- Documentum
- Domino
- Exchange Server
- IBM DB2
- Informatica
- Internet (HTTP)
- Microsoft SQL
- iManage Server
- Lotus Quickplace
- Sybase
- eRoom
- Stemming
- FileNet
- Translation (WBC) as a service
- Flexible Query Manipulation
- Enterprise Search
- World Decomposition
- Flexible File System
- N-grid indexing
- MySQL
- Vignette
- ODBC Databases
- OpenText LiveLink
- PCLDoc
- Informik
- Microsoft SharePoint
- Moreover
- Enterprise Architecture
- NIJ File System
- Network File Systems
- Site Under Notegrity
- Lexis Nexus
- WebSphere.....



IDOL elemi funkciók (500+)

- 2D/ 3D Clustering
- Auto Query Guidance
- Acoustic Signature
- Active Matching
- Agents
- Alerting
- Auto Language Detection
- Boolean & Legacy Operations
- Breaking News Clustering
- Categorization
- Collaboration
- Community
- Concept Highlighting
- Concept-Query Summarization
- Conceptual Retrieval
- Context Summarization
- Cross-Modal Suggest
- Dynamic n-dimensional Taxonomy Generation
- Dynamic XML Consumption
- Education
- Exact Phrase Matching
- Expertise Location
- Explicit Profiling
- Face Recognition
- Field Modulation
- Frame Analysis
- Fuzzy Matching
- Hot Clustering
- Hyperlinking
- Image Analysis
- Image Association
- Implicit Profiling
- Keyword Search
- Mail Object Identification
- Melody Classification
- Melody Identification
- Metadata Recognition
- Natural Language Retrieval
- Object Identification
- Object Recognition
- Ontology Generation
- Parametric Refinement
- Phrase Spotting
- Proper Name Identification
- Query by Example
- Real-time Aggregation
- Routing
- Scene Detection
- Script Alignment
- Soundex Matching
- Speaker Identification
- Speaker Recognition
- Spectrographic Analysis
- Speech Recognition
- Spell Checking
- Tag Reconciliation
- Transcription
- Video Analysis
- Voice Printing
- Word Spotting
- Work Groups
- XML Tagging...



Autonomy termékcsalád

	POWER	PROTECT	PROMOTE
Terület	Meaning Based Computing	Meaning Based Governance	Meaning Based Marketing
Megoldások	Search BPM OEMs	eDiscovery Archiving & Records Content Mgmt Compliance	Web Content eBusiness Marketing Optimization Rich Media Mgmt Contact Center
Platform	Autonomy Intelligent Data Operating Layer (IDOL)		



Potenciális lehetőségek...

- Az információ 100%-ának felhasználása
- Felismerni a vásárlási szándékot és reagálni rá
- Termékeladások valós idejű követése
- Hitelminősítés



Konkrét alkalmazások

- Keresés
 - Állambiztonság
 - Amerikai jogi cégek
 - Hírportálok
 - Marketing tevékenység támogatása
- Dokumentum menedzsment
 - Bankok, gyógyszeripari cégek: jogszabályi megfelelés
- Video analízis
 - Biztonsági megoldások: helyzet felismerés, arcfelismerés
 - Nyomkövetés (tárgyak, személyek)



Kérdések...

- ...válaszok (?)



Ez itt a reklám helye...

- Önálló labor
- Diplomatervezés



Köszönöm a figyelmet!

Wagner Kornél
kornel.wagner@hp.com



Adatbázisok haladóknak 2012.

2012. október 2.