



**Передовые решения HP для
работы с неструктурированными
данными**

Евгений Олейник

Руководитель департамента
Autonomy IM в Центральной и
Восточной Европе

POWER PROTECT PROMOTE

Что такое Autonomy?

Подразделение HP с ресурсами по всему миру

Основана в 1996 учеными из университета Кембридж

До поглощения HP – одна из самых быстрорастущих ИТ компаний за всю историю

Приобретена HP в Октябре 2011 за \$11 млрд

Основной движок IDOL – 170 патентов, 1000+ типов файлов, 500+ функций, 400+ источников, 150 языков

Лидер в:

Обработке неструктурированных данных воспринимаемых человеком:

- Текст
- Аудио
- Видео

Облачных вычислениях

Обработке больших данных и аналитике

Защите данных

Маркетинговой оптимизации

Key stats:

65,000+ заказчиков

Более 50 PB данных хранимых в облаке; Рост подписчиков облачных услуг: +41%

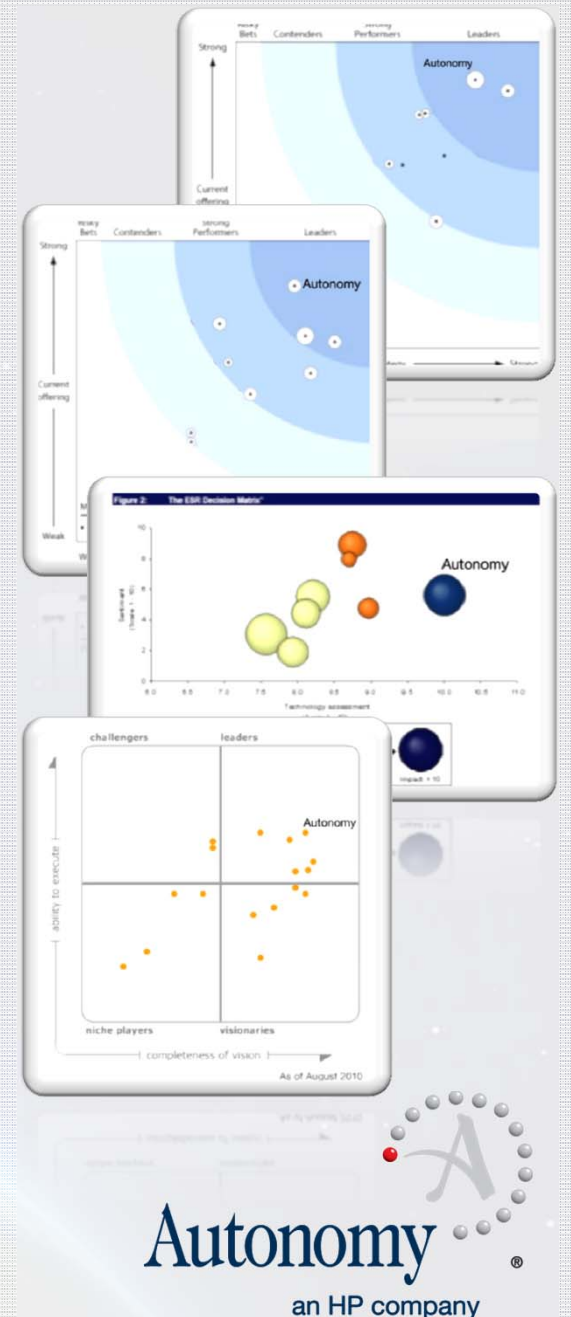
10 из 10 крупнейших глобальных банков

19 из Financial Times топ 25

10 из 10 крупнейших глобальных аудиторских компаний

Движок для более чем 20,000 сайтов, включая 9 из 10 топ брендов

НИОКР: \$200млн



Более 65,000 заказчиков



Media

Reuters
Bloomberg
LexisNexis
CNN
Knight Ridder
Martha Stewart Omni.
Wolters Kluwer
BBC
BskyB
National Public Radio
McGraw Hill
Reed Elsevier
HBO
Forbes
Associated Newspapers
Belga NV
Shanghai News Agency
Rai
Channel 4
Condé Nast
HarperCollins

Energy

Norsk Hydro
BP
US Dept of Energy
Halliburton
GE Energy Services
SPE.org
MOL
Encana
Devon Energy
CPS Energy

IT

IBM
Logica
OpenMarket
Seagate
Oracle
Symantec
Finsiel
Cebit
OTEnet
Juniper Networks
Wipro
Accenture
HP
Sun
Sybase
Unisys
Cap Gemini
EMC Corporation
Cisco Systems
McAfee

E-Commerce

Tesco
RadioShack
BUT
Safeway.com
Napster
Toys R US
The NewsMarket
Albertsons
AVIS

Government/ Investig.

US Dept. Of Defense
US Dept. Of Health
NASA
UK Police Forces
US Army
DERA
Swiss Army
Port of Singapore
Customs & Excise
UK Cabinet Office
UK COI
UK MOD

Pharmaceutical

Astra Zeneca
Novartis
GlaxoSmithkline
Organon
SAIC
Saludalia
LocusMedicus
Pfizer
Roche
Monsanto

Legal

Freshfields
White & Case
Butterworths
Law Society
Addleshaw
Linklaters
Allen & Overy
Slaughter & May
Baker & McKenzie

Telecommunications

T-Mobile
Alcatel-Lucent
Sprint
Qwest
MCI
Sonera
AT&T
Telia
BT
Telecom Italia
3
Vodafone

Manufacturing/ FMCG

General Motors
BAE Systems
Nestle
Volkswagen
Chrysler
Skanska
Agilent
Philip Morris
Ford Motor Company
Goodyear
Aluminium Pechiney
Boeing
Lexmark
Texas Instruments
Agilent Technologies
DTI
Arup
Frito-Lay
Kohler
Eastman Kodak
John Deere
Mercedes-Benz
Nissan
Sony

Financial

Royal SunAlliance
Dresdner Kleinwort
Fidelity
Deutsche Bank
Lloyds TSB
Danske Bank
Bank of Montreal
New York Life
ABN Amro
AMP Henderson
Darier Hentsch
Goldman Sachs
Banco de Espana
Schroders
UBS
Fireman's Fund
Insurance
Merrill Lynch
Barclays
Credit Suisse
Commerzbank
Legal and General
Abbey
JPMorganChase

Professional Services

KPMG
PWC
AMR Research
Butler Group
Choicepoint

Включая....

Autonomy
an HP company

*400 OEM поставщиков
интегрируют Autonomy:*

XEROX® CISCO

Adobe

Novell®
VERDASYS.

EMC²
where information lives

ORACLE®

KANA

UGS
The PLM Company

SYBASE®

CITRIX®

IBM

MatrixOne

TUMBLEWEED
COMMUNICATIONS

UNISYS

OPENWAVE®

*Используют Autonomy как
платформу:*

BBC

Bloomberg

devon

Schneider
Electric

AstraZeneca

BT
HYDRO

FedEx®

Schlumberger

BDO

ERICSSON

gsk
GlaxoSmithKline

Ford

3

NISSAN

U.S. DEPARTMENT OF
HOME LAND SECURITY

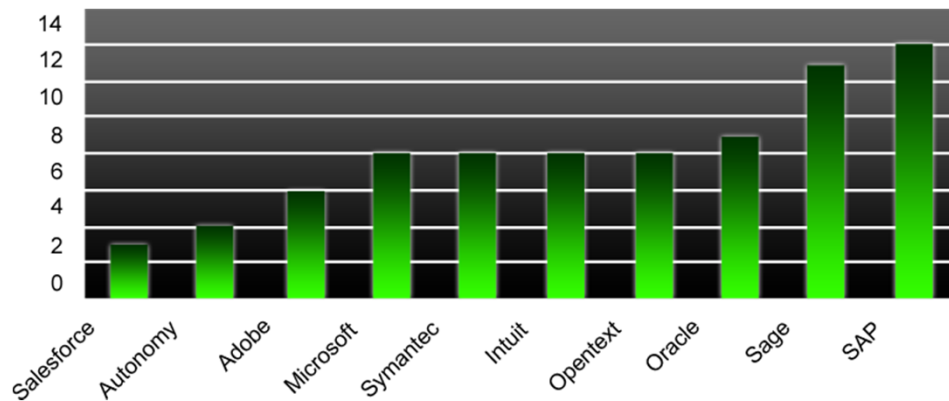
Shell

Лидер в обработке больших данных

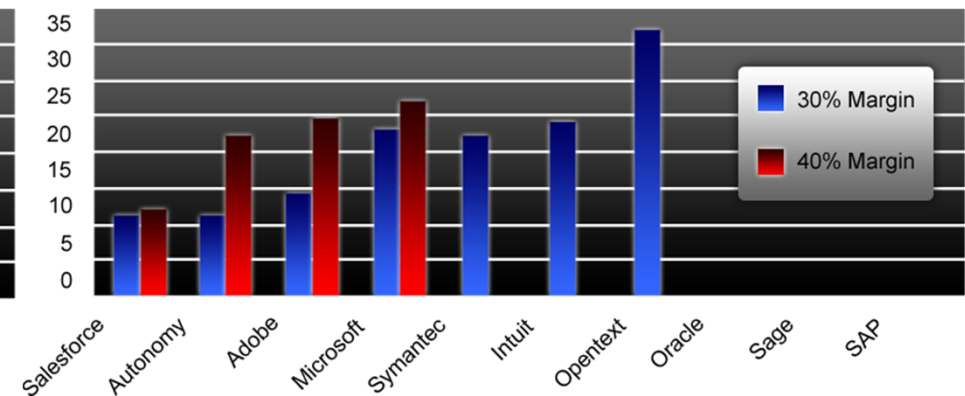


- «Вычисления со Смыслом» меняют подход человека к обработке информации
- Autonomy понимает 100% всей информации – и структурированной и неструктурированной
- Autonomy - лидер по росту среди компаний разработчиков ПО

Years to \$50m in Revenues



Years to 30% and 40% Operating Margin



Области специализации Autonomy



Autonomy®

Преимущество	<i>Эффективность</i>	<i>Безопасность</i>	<i>Развитие</i>
Наше отличие от других	Вычисления со смыслом*	Управление со смыслом	Маркетинг со смыслом
Основные Специализации	Поиск Процессы Интеграция Аналитика	eDiscovery Архивирование и управление док-тами Управление контентом Контроль соблюдения правил и политик Бекап /восстановление	Управление веб-контентом Электронная коммерция Оптимизация Управление медиа контентом Колл-центры
Платформа	Autonomy Intelligent Data Operating Layer (IDOL)		

* Meaning-based computing – выделение смысловых единиц

Мнения независимых аналитиков



Autonomy is essential, and it's attractive

There are lots of reasons why investors should buy shares in Autonomy.



Autonomy won the enterprise search wars



October 2008

Autonomy Fastest Growing Full Service EDD Supplier Among the AM Law 200



Gartner

"the next quantum leap in productivity will come from the use of IT systems that analyse structured and unstructured data sitting within organisations, searching and analysis of data from all sources will allow businesses to optimise the information about customers and product sets. It will contribute to an improvement in all aspects of business operations. The British company Autonomy is at the cutting edge of the new search technology." He believes this sector of the IT industry will ultimately earn multi-trillion dollar revenues.

Peter Sondergaard, Head of Worldwide Research at Gartner

ComputerWeekly

Autonomy dominates search space, reports IDC

Autonomy sparkles in a class of its own

See [Autonomy: All sides complete](#)



Fastest growing and largest market share by far

A

Call...
Crest...
...can...
...clan...
...stare...
...The...
...maly...
...can...
...to's...
...Dey...
...ing...
...as...
...the...
...dress...

Name: Smith

Address: 46 Acacia avenue

Age: 32

Code: 00403034

Reference FFRRF55434





Email
Documents
Apps
Web
Social Media
Desktop
Content Management
Phone Calls
Customers
TV Shows
Video
XML
Catalogs
Instant Messenger
Rich Media
Print

Enterprise Search
Portals
Call Center
Email Archiving
Compliance
Real-Time Policy
Content Aggregation
CRM
Risk Management
Video Search
Reputation Analysis
Publishing
Alerting
Marketing Optimization
eCommerce
Customer Interaction
Multichannel
Dynamic Web Delivery
Rich Internet Applications
Manage in Place

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 +

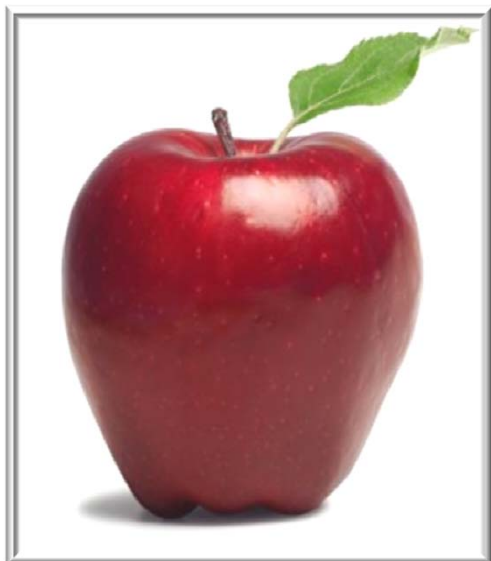
Keyword Search



- Определение идей, а не только ключевых слов
- Открытая архитектура SOA ACI XML
- Одна платформа для всех функций
- Одна платформа обрабатывает все типы данных: аудио, видео, неструктурированные, XML, структурированные и т.д.
- Технологии новейшего поколения: сложные запросы, поиск внутри аудио и видео, кластерный анализ...

***“Autonomy is the market leader in providing software to automate the classification and management of data.”
SVM, 2010***

$$H = -\sum p_i \cdot \log_2(p_i) \quad P(\theta | x) = \frac{P(x | \theta) \cdot P(\theta)}{\sum_{\theta' \in \Theta} P(x | \theta') \cdot P(\theta')}$$



- **Информация может иметь различные формы**
Текст, Звук, XML, Видео и Аудио
- **Но также важен контекст**
“Яблоко это фрукт?”
“ИТ компания?”
*“Глазное **яблоко**”*
“Яблоко раздора...”
*“партия **Яблоко**”*
- **Значение всегда многогранно**
- **Но значение связано с контекстом**
- **Не совпадающие значения всегда отличаются по контексту**

*Ключевые слова и метаданные не решают эту задачу.
Для работы с неструктурированными данными нам нужна система автоматизированной обработки и извлечение информации, по аналогии с человеческим восприятием.*

Понимание смысла и автоматизация – долой ручной труд!



Автоматизация процессов

- Notes
- Database
- Document Management
- News Feeds
- Email
- Internet
- Files
- XML
- Audio/ Media

$$H = -\sum p_i \cdot \log_2(p_i)$$


**Теория информации
Шаннона и формула
Байеса**

$$P(\theta | x) = \frac{P(x | \theta) \cdot P(\theta)}{\sum_{\theta' \in \Theta} P(x | \theta') \cdot P(\theta')}$$

Integration
Through
Understanding

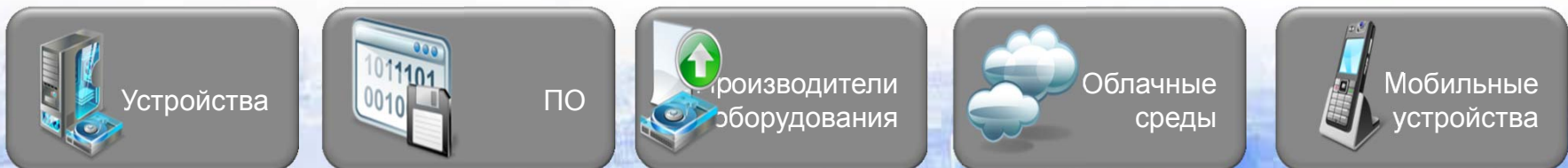
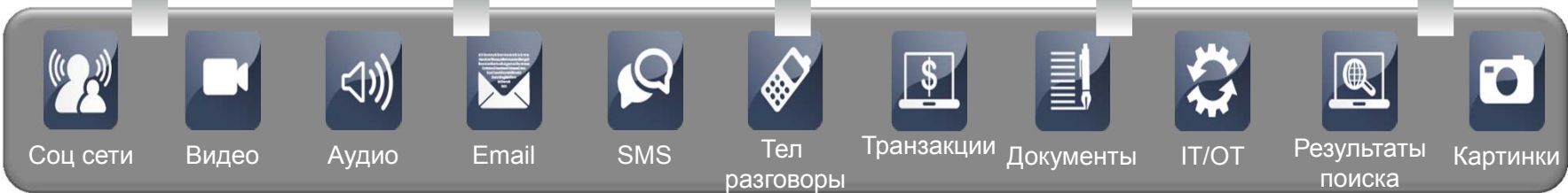
- Aggregation
- Automatic categorization
- Авто-создание гиперссылок
- Профилирование
- Персонализация
- Командная работа
- Предоставление контента
- Выемка информации
- Перенаправление
- Уведомления о новом контенте

Платформа IDOL



Современная платформа IDOL10™

Более 400 коннекторов



Over 500 IDOL Functions

2D/ 3D Clustering
Acoustic Signature
Active Matching
Agents
Alerting
Auto Language Detection
Auto Query Guidance
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...

Over 400 Connectors

Centera	iManage	ODBC
Documentum	IBM Content Manager	Omni
Cognos	Java	Oracle
DM5	KVS	PC Docs
eRoom	LiveLink	POP3
Exchange	Lotus Notes	RSS
Fatwire	MatrixOne	SAP
File System	Meridio	SharePoint
FileNet	Moreover	Siebel
FTP	MS Access (MDB)	Stellent
Groove	Netware	Webservice
HTTP	NNTP	



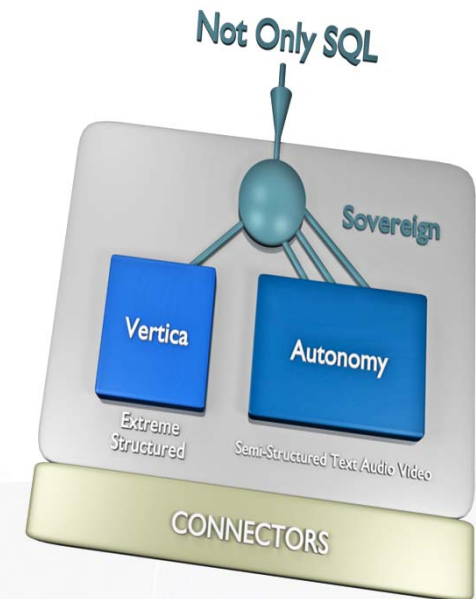
POWER PROTECT PROMOTE



Autonomy[®]
an HP company

Демонстрация

- Autonomy – лидер в понимании смысла информации воспринимаемой человеком
- IDOL – единая платформа для всех решений Autonomy
- Проверенные в крупнейших компаниях мира решения
- Наивысшая эффективность за счет автоматизации
- **IT индустрия может справиться с 15% проблем, мы можем все 100%**





Спасибо!
oleynik@hp.com