

KAM SMĚŘUJE VYHLEDÁVÁNÍ ...

Pavel Kocourek
SEARCH TECHNOLOGIES, formerly INCAD
INFORUM 2017 – 31. 5. 2017

20 MINUT

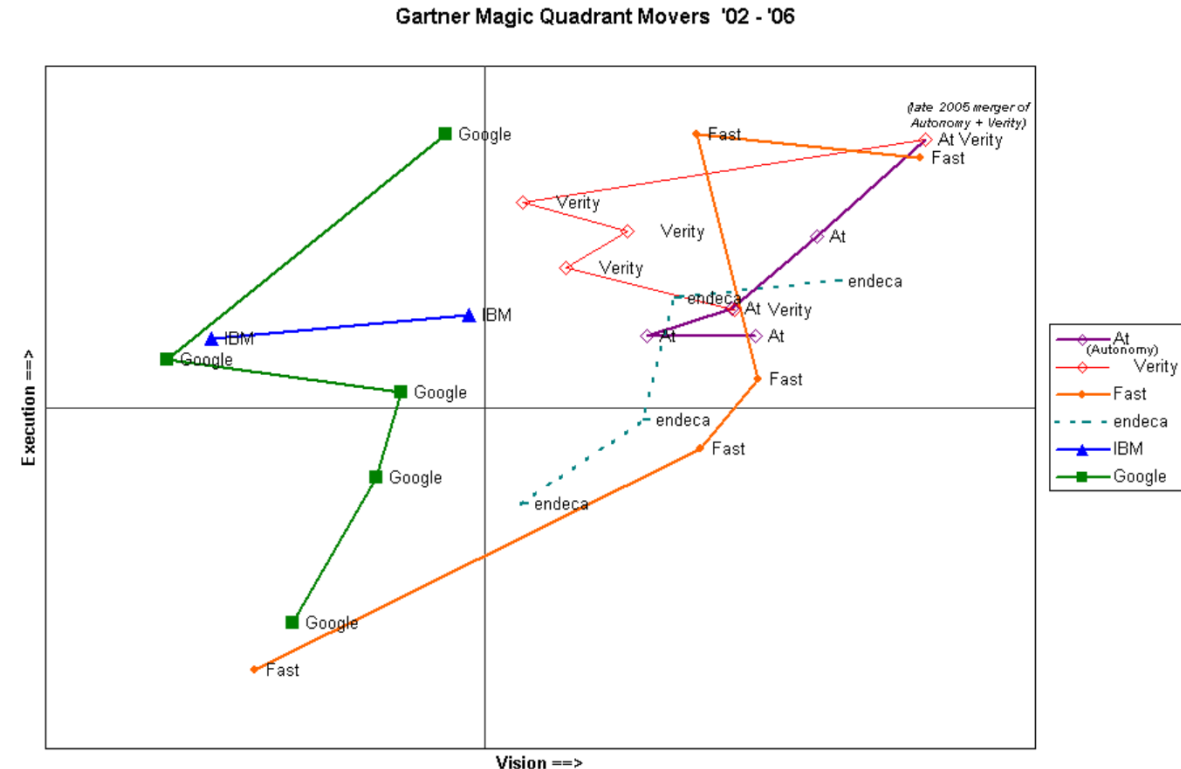
- Tradiční Enterprise Search a OpenSource
- Lucene / SOLR vs Elastic
- Nová generace vyhledávání ... ?
- Kam se „rozbíhá“ další vývoj ?

TRADIČNÍ VYHLEDÁVÁNÍ

MALÉ OHLÉDNUTÍ ZA ES

Už je to 10 let

- FAST > Microsoft
- Endeca > Oracle
- Verity > Autonomy > HP
- Vivisimo > IBM
- Exalead > Dassault Systemes
- ...



Open source nástroje, v podstatě převzaly tradiční vyhledávání.

Komerční dodavatelé směřují k „nové generaci“ vyhledávačů ...

OPEN SOURCE - SOLR NEBO ELASTIC ?

SOLR vs ELASTIC

- Který je lepší SOLR nebo ELASTIC?
 - Který je rychlejší?
 - Které řešení je lépe škálovatelné?
 - Který umí A, B, a C?
 - Který je jednodušší na správu?
-
- Který máme nasadit? Proč a co doporučujete?

OPEN SOURCE



Apache Lucene™ je vysoce výkonná, plnohodnotná knihovna vyhledávacího systému napsaná v jazyce Java. Technologie je využitelná pro téměř libovolnou aplikaci, která vyžaduje fulltextové vyhledávání, bez závislosti na platformě.
1999 Doug Cutting > 2010 Lucene/SOLR > 2017 r 6.5.1



Solr je vysoce spolehlivý, škálovatelný a odolný vyhledávací systém. Poskytuje distribuovanou indexaci, replikace, rozložení výkonu dotazování, fail-over a recovery, centralizovanou konfiguraci a další funkce.
2004 Yonik Seeley > 2010 SOLR/Lucene > 2017 r 6.5.1



Elasticsearch je distribuovaný vyhledávací a analytický nástroj s RESTful rozhraním. Systém je použitelný na širokou řadu úloh společně s nástroji z Elastic Stack.
2010 Shay Banon > 2017 r 5.4.0

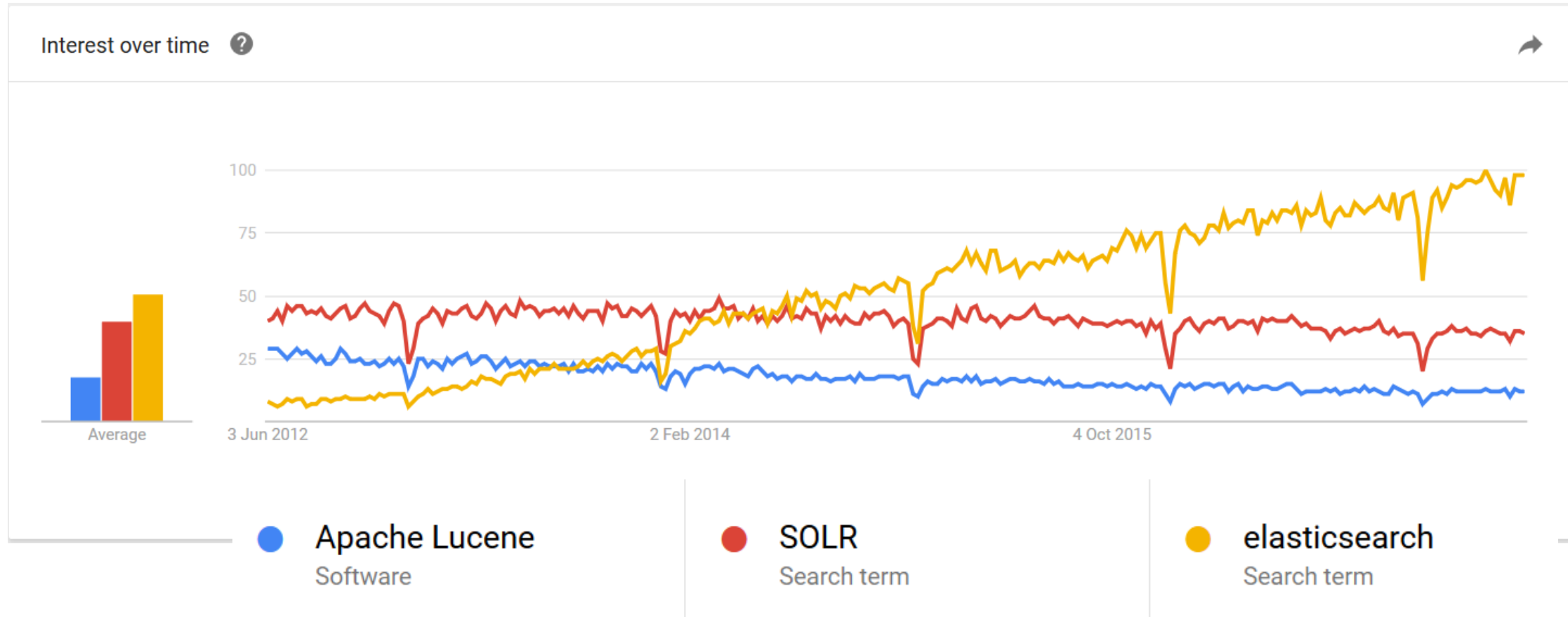
SOLR

- Díky své vyspělosti, širokému nasazení a aktivní komunitě je vyhledávací systém Lucene/SOLR upřednostňován zkušenými vývojáři a administrátory
- SOLR poskytuje velmi detailní funkční prostředí, včetně detailní dokumentace
- K dispozici je široká řada plug-inů
- Lucene/SOLR má rozsáhlou komunitu vývojářů

ELASTIC

- Mladší, ale již velmi rozšířený vyhledávací nástroj vyvíjený s větším zaměřením na získávání informací z dat obecně, nikoliv pouze na vyhledávání.
- Součást "ELK stack" (Elasticsearch, Logstash, Kibana)
- Elasticsearch umožňuje snadněji realizovat velká vyhledávací řešení
- RESTfull search, umožňuje schema-free nasazení a je orientovaný na dokument
- Navržen s ohledem na využití v cloud prostředí
- Společnost tvrdí, že od 2014 má více než milion stažení měsíčně

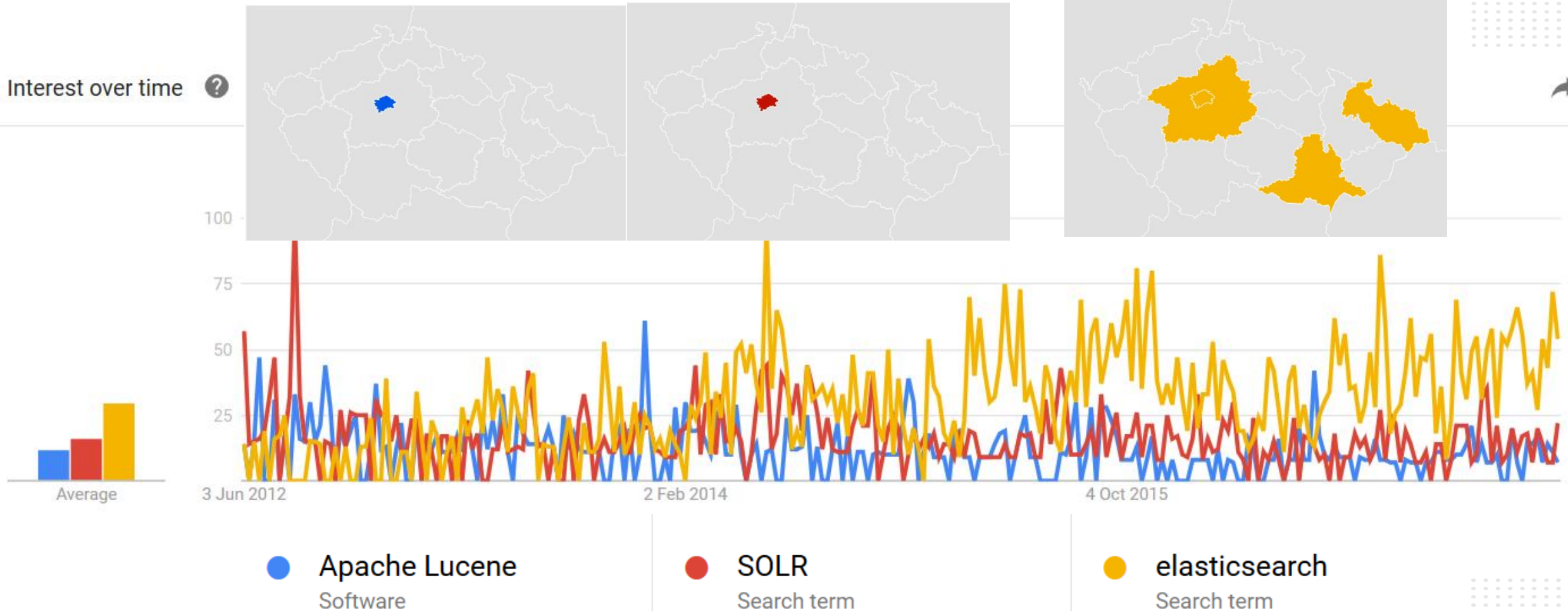
OPEN SOURCE - popularita



<https://trends.google.com/trends/explore?q=%2Fm%2F02l7pz,SOLR,elasticsearch>

OPEN SOURCE – popularita CZ

Interest over time ?



<https://trends.google.com/trends/explore?q=%2Fm%2F02I7pz,SOLR,elasticsearch>

OPEN SOURCE

	Solr	Elasticsearch
Installation and Configuration	Supported by detailed documentation	More intuitive
Indexing/Searching	Text-oriented	Better performance of analytical queries
Scalability and Clustering	Provides SolrCloud	Better inherent scalability and designed for the cloud
Community	A much bigger ecosystem of community	A growing community though not a complete open source mindset
Documentation	Very well-documented	Lacks in documentation

<https://logz.io/blog/solr-vs-elasticsearch/>

DETAILY ? ... NA WEBU

Apache Solr vs Elasticsearch

The Feature Smackdown

API

Feature	Solr 6.2.1	ElasticSearch 5.0
Format	XML, CSV, JSON	JSON
HTTP REST API	✓	✓
Binary API	✓ SolrJ	✓ TransportClient, Thrift (through a plugin)
JMX support	✓	✗ ES specific stats are exposed through the REST API
Official client libraries	Java	Java, Groovy, PHP, Ruby, Perl, Python, .NET, Javascript Official list of clients
Community client libraries	PHP, Ruby, Perl, Scala, Python, .NET, Javascript, Go, Erlang, Clojure	Clojure, Cold Fusion, Erlang, Go, Groovy, Haskell, Java, JavaScript, .NET, OCaml, Perl, PHP, Python, R, Ruby, Scala, Smalltalk, Vert.x Complete list
3rd-party product integration (open-source)	Drupal, Magento, Django, ColdFusion, Wordpress, OpenCMS, Plone, Typo3, ez Publish, Symfony2, Riak (via Yokozuna)	Drupal, Django, Symfony2, Wordpress, CouchBase
3rd-party product integration (commercial)	DataStax Enterprise Search, Cloudera Search, Hortonworks Data Platform, MapR	SearchBlox, Hortonworks Data Platform, MapR etc Complete list
Output	JSON, XML, PHP, Python, Ruby, CSV, Velocity, XSLT, native Java	JSON, XML/HTML (via plugin)

Infrastructure

Feature	Solr 6.2.1	ElasticSearch 5.0
Master-slave replication	✓ Only in non-SolrCloud. In SolrCloud, behaves identically to ES.	✗ Not an issue because shards are replicated across nodes.
Integrated snapshot and restore	Filesystem	Filesystem, AWS Cloud Plugin for S3 repositories, HDFS Plugin for Hadoop environments, Azure Cloud Plugin for Azure storage repositories

Indexing

Feature	Solr 6.2.1	ElasticSearch 5.0
Data Import	DataImportHandler - JDBC, CSV, XML, Tika, URL, Flat File	[DEPRECATED in 2.x] Rivers modules - ActiveMQ, Amazon SQS, CouchDB, Dropbox, DynamoDB, FileSystem, Git, GitHub, Hazelcast, JDBC, JMS, Kafka, LDAP, MongoDB, neo4j, OAI, RabbitMQ, Redis, RSS, Sofa, Solr, St9, Subversion, Twitter, Wikipedia
ID field for updates and deduplication	✓	✓
DocValues	✓	✓
Partial Doc Updates	✓ with stored fields	✓ with _source field
Custom Analyzers and Tokenizers	✓	✓
Per-field analyzer chain	✓	✓
Per-doc/query analyzer chain	✗	✓
Index-time synonyms	✓	✓ Supports Solr and Wordnet synonym format
Query-time synonyms	✓ especially via non-lucene-synonyms	✗ Technically, yes, but practically no because multi-word/phrase query-time synonyms are not supported. See this

<http://solr-vs-elasticsearch.com/>

NOVÁ GENERACE ...

ENTERPRISE SEARCH 2017

- **Machine Learning transformuje ES do Intelligent Search**
- **Pokrok v oblasti Cyber Security přinese nové možnosti ... zejména kvůli cloud technologiím**
- **Real-Time Informace budou mnohem dostupnější “S ohledem na machine learning”**
- **Doporučování bude „chytřejší“: Proactive search**
- **Vyhledávání změní digitální pracovní prostor**

INSIGHT ENGINES 2017

- Pracující s přirozeným jazykem
 - poskytují přirozenější přístup k informacím postupy, které podnikové vyhledávání nemělo
- Komplexní
- Proaktivní
 - aplikují metody hodnocení relevance pro popis, objevování a organizování a analýzu dat.



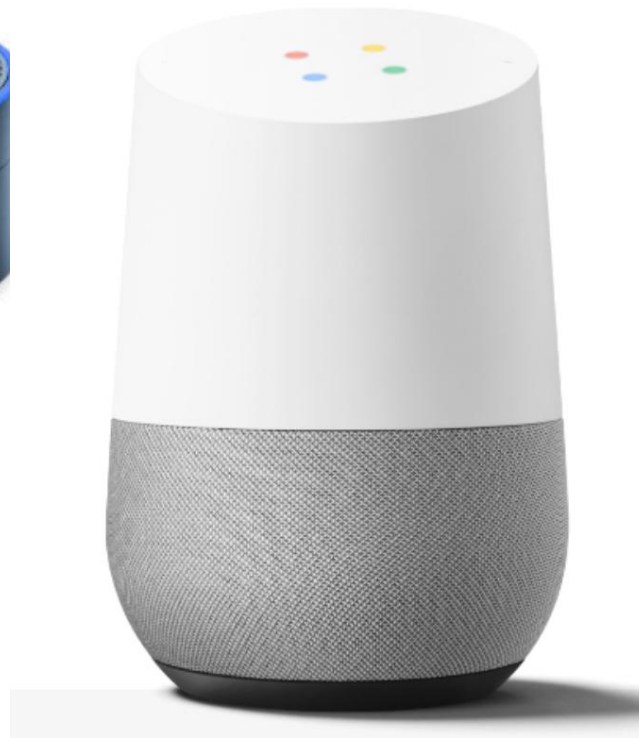
BUSINESS INTELLIGENCE AND ANALYTICS PLATFORMS 2017

Search a Analytics - konvergence
Cloud a opensource - mainstream
NLP a hlasové vyhledávání
IoT



ZA ČÍM JE VYHLEDÁVÁNÍ SCHOVANÉ

...



INTELIGENTNÍ DIGITÁLNÍ ASISTENTI/KY



Úterý 30. května 2017. Ferdina

iDNES.cz > Zprávy | Kraje | Sport | Kultura | Ekonomika | Bydlení | **Technet** | Ona | Revue | Auto | ☰ Další

Audio | Video | Tv | Foto | PC & Mac | Software | Notebooky | Web | **Věda & Vesmír** | Vojenství | Autoři | Windows 10



Když se DLC vymknou z rukou. Virtuální obsah za desítky tisíc korun



GLOSA: Je načase přestat, uhádla Mordparta II, když dokončila čistku



Fotili jsme v plavkách nemám za

Velký test digitálních asistentek: překvapily inteligencí i neschopností

13. dubna 2017, aktualizováno 10:41

Digitální asistentky do chytré domácnosti, tak se prezentují dvě zařízení dostupná na americkém trhu. Poslouchají každé vaše slovo a slibují, že se vám pokusí pomoci. Podrobně jsme otestovali jejich deklarované dovednosti. Posuďte sami, zda tudy vede cesta k oné chytré domácnosti budoucnosti.



http://technet.idnes.cz/test-digitalni-asistentky-amazon-echo-alexa-google-home-pjy-/tec_technika.aspx?c=A170410_074844_tec_technika_pka

INTELIGENTNÍ DIGITÁLNÍ ASISTENTI/KY

Amazon **Alexa**

Apple **Siri**

Microsoft **Cortana**

Google assistant – nepojmenovaný 😊

Počítačový program, který pracuje jako „intelligent personal assistant“ případně „knowledge navigator“.

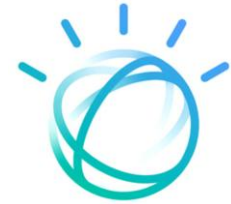
Využívá NLP pro zodpovězení dotazu, doporučování a provedení akcí delegováním požadavků na webové služby.



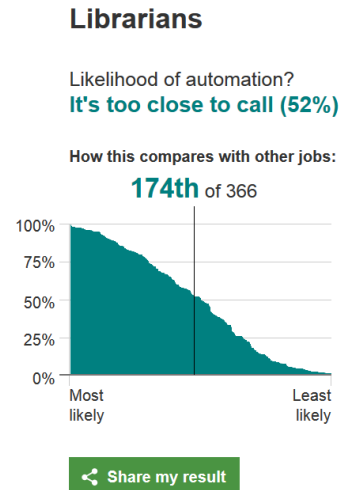
QA

IBM Watson

Výpočetní systém pro zodpovídání dotazů (QA) vytvořený IBM pro aplikaci Zpracování přirozeného jazyka NLP, vyhledávání informací, reprezentaci znalostí, automatické zdůvodňování a učící se technologie v otevřené doméně.



JAK TOHLE KNIHOVNY?



<http://time.com/4742543/robots-jobs-machines-work/>
<http://www.bbc.com/news/technology-34066941>



Will a Robot Take My Job?

Enter your occupation



You might be safe

Librarians

43% of your job can
be done by a robot

Your job is safer than
47.8% of jobs

TASKS A ROBOT CAN DO

- Order instructional or library materials or equipment
- Maintain inventories of materials, equipment, or products
- Process library materials
- Maintain operational records
- Write grant proposals
- Construct exhibits or parts of exhibits
- Select educational materials or equipment
- Inspect materials or equipment to determine need for repair or replacement
- Search information sources to find specific data

TASKS A ROBOT CAN'T DO

- Classify materials according to standard systems
- Advise educators on curricula, instructional methods, or policies
- Direct department activities
- Serve on institutional or departmental committees
- Direct activities of subordinates
- Teach others to use technology or equipment
- Train staff members
- Negotiate purchases or contracts
- Promote educational institutions or programs
- Help patrons use library or archival resources
- Confer with others to conduct or arrange operational activities
- Write articles, books or other original materials in area of expertise
- Develop instructional materials
- Organize informational materials
- Plan community programs or activities for the general public
- Develop policies or procedures for archives, museums or libraries
- Attend training sessions or professional meetings to develop or maintain professional knowledge
- Develop library or archival databases
- Compile specialized bibliographies or lists of materials

Source: McKinsey&Company

DO MRAKŮ ... MOBILNĚ ... LOKÁLNĚ ... PERSONALIZOVANĚ ...

CO PŘIJDE ? 😊

Co přinesou technologie zhruba víme ...

Nejtěžší je odhadnout, jak technologie změní lidské chování



Source: Gartner (July 2016)

Find better Answers
www.searchtechnologies.com/cz