MINISTERSTVO PRO MÍSTNÍ ROZVOJ
Národní orgán pro koordinaci

Data, vizualizace a infografiky pro podporu core aktivit TA ČR

Systémy, principy a výstupy

Martin Víta, 3. 10. 2018

EVROPSKÁ UNIE
Fond soudržnosti
Operační program Technická pomoc

MINISTERSTVO
PRO MÍSTNÍ
ROZVOJ ČR

T A

Č R

Výzkum užitečný pro společnost

T A
Č R

Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Data, vizualizace a infografiky pro podporu core aktivit TA ČR

Systémy, principy a výstupy

3. 10. 2018

Technologická agentura České republiky

Martin Víta

Metodik QC/QA - Projekt ProEval

T A
Č R

Pohled skrze data na…

- **Monitoring projektů / programů**
- **Evaluace programů**
- Zkvalitňování veřejných soutěží
- Ad hoc analýzy pro potřeby TA ČR a dalších institucí
- …

Existence relevantních datových zdrojů + trend evidence based policy: ideální kombinace

Data ve výzkumu, vývoji a inovacích v ČR

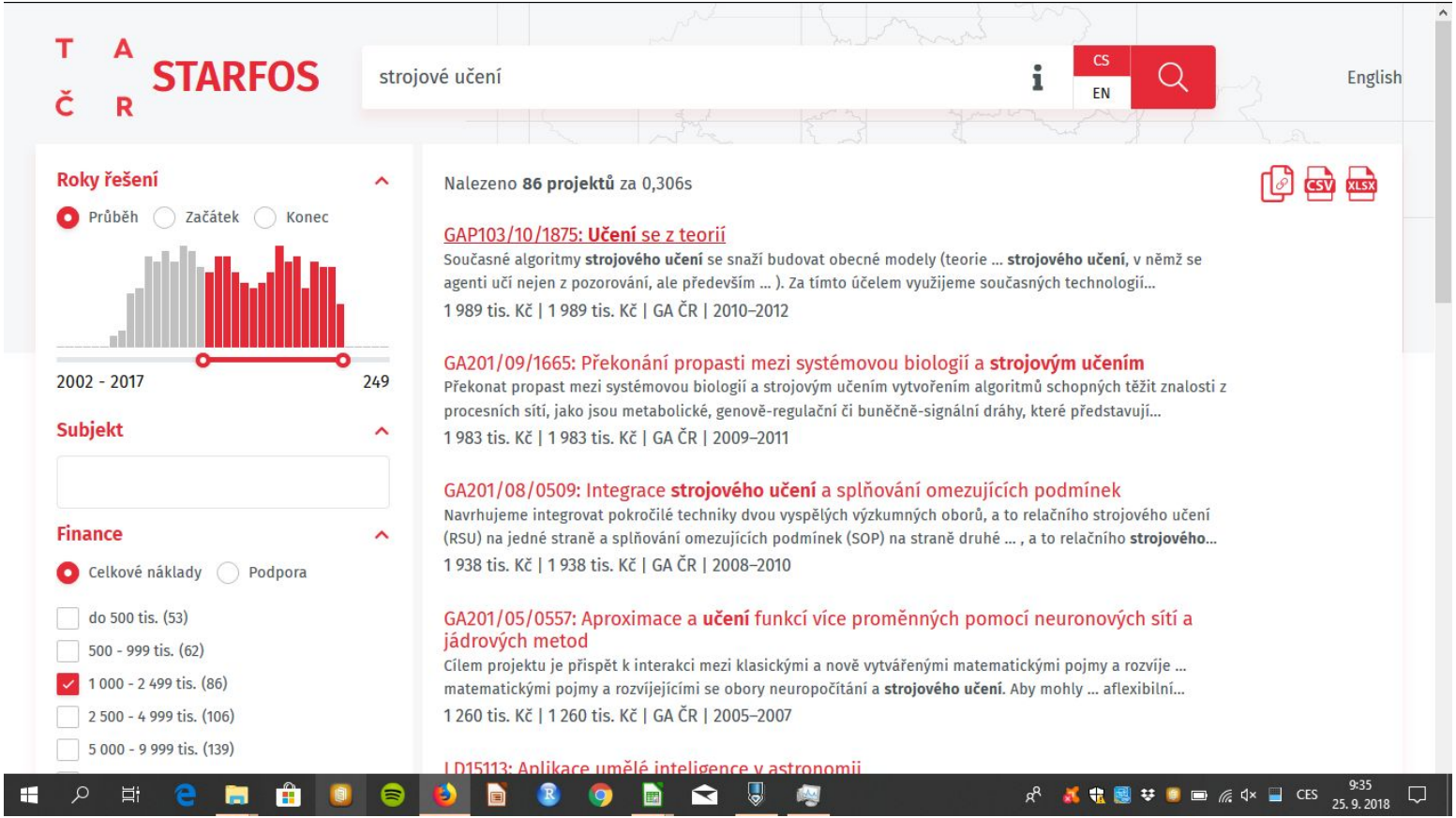
- **Klíčová role IS VaVaI** (dnes na rvvi.cz)
  - problém: omezená funkcionalita týkající se vyhledávání/filtrování
- **Interní data TA ČR**
- Oficiální statistika ČSÚ a EUROSTAT
- Ostatní zdroje (ČR a EU)

Obecný problém integrace datových zdrojů

Systém STARFOS jakožto nástroj na integraci dat, interaktivní filtrování a vyhledávání

T A
Č R

STARFOS – ukázka



Pilíře systému STARFOS - od dat k reportům

- **Datová základna** (na počátku data rvvi.cz)
- **Fasetová klasifikace** (a vyhledávání, resp. filtrování)
  - *systém třídění umožňující zařazení objektu do více tříd (výzkumná organizace / příjemce podpory od daného poskytovatele / instituce z daného kraje / …)*
  - *obvyklý např. ve velkých e-shopech*
- **Search engine** (solr)

T A
Č R

Vize systému STARFOS

- Systém **integrující různorodé relevantní datové zdroje**
- Komfortní aplikace na **vyhledávání a filtrování** entit různých typů ve VaVaI (instituce, výsledky, projekty …)
- Nástroj na tvorbu **interaktivních reportů** – inspirace D3 (data driven documents)

Část II – vizualizace

www.tacr.cz

Od reportů k vizualizacím

- Motivace – jednoduché otázky
  - Jak získat přehled o zaměření projektů daného programu v “kompaktní formě”?
  - Které instituce jsou z pohledu daného programu “významné?” - viz další část přednášky
- Nutnost přejít od reportů k vizualizacím
- *Ukázka (shlukování na množině projektů: separátní PDF)*

T A
Č R

Principy práce s entitami reprezentovanými textově

- Předpokládáme, že entity máme popsány pomocí **textových dokumentů** (např. instituce seznamy klíčových slov z jejich projektů)
- **Vektorová reprezentace** textů
- Výpočet jejich **podobnosti** (vzdálenosti ve vektorovém prostoru)
- Analýza **grafu podobnosti** (např. detekce komunit/clusterů)

T A
Č R

Use case – analýza helpdesku

- Využití popsaného přístupu **v jiném kontextu**
- Prakticky motivovaná otázka: “Jaké jsou typické tématické okruhy dotazů uchazečů?”
- *Ukázky vizualizací (separátní soubory):*
  - *shluky/komunity otázek*
  - *wordcloudy nejčastějších slov/slovních spojení*

T A
Č R

Část III – analýza grafových dat

www.tacr.cz

T A
Č R

Analýza grafů – př. se souvýskyty

- Otázka souvýskytů určitých entit, např. klasifikačních tříd oborů (CEP)
  - *hlavní a vedlejší obor projektu*
- Výstup – opět graf podobnosti
- *Ukázka (separátní PDF)*

Analýza grafů – grafy spolupráce

- Motivační otázky:
  - Které instituce jsou z pohledu daného programu “významné”?
  - Vznikají v rámci programu kliky či komunity spolupracujících institucí (bez apriorní negativní konotace)
- *Ukázka grafu (separátní PDF)*

Modelování významnosti entit – centrality

- Hlavní myšlenky jednotlivých typů centralit:
  - degree centrality
  - betweenness centrality
  - closeness centrality
  - triangle
  - **eigenvector centrality**

Instituce s nejvyšší eigenvector centrality v první VS programu ALFA

Instituce	Eigenvector centrality
České vysoké učení technické v Praze	10.0
Vysoké učení technické v Brně	0.53
Vysoká škola báňská - Technická univerzita Ostrava	0.32
Centrum dopravního výzkumu, v.v.i.	0.26
Ústav teorie informace a automatizace AV ČR, v. v. i.	0.21
Západočeská univerzita v Plzni	0.20
EVEKTOR, spol. s r.o.	0.19
HVM PLASMA, spol. s r.o.	0.19
VARS BRNO a.s.	0.18

T A
Č R

Část IV – TA ČR a infografiky

www.tacr.cz

T A
Č R

Od vizualizací k infografikám

- Vizualizace jakožto vhodně zvolená grafická reprezentace konkrétních dat
- Infografika jakožto souhrn klíčových informací o dané problematice grafickou a interaktivní formou
- *Procházka infografikami* http://visual.tacr.cz

T A
Č R

Část V – datové zdroje, TA ČR a otevřená data

www.tacr.cz

Datové zdroje TA ČR

- **INKA** – mapování inovačních kapacit
  - https://inkaviz.tacr.cz (prohlížeč dat)
- **TA ČR v číslech** - statické prostředí, které nás navedlo k visualu
  …další zdroje na https://tacr.cz/index.php/cz/test-programy/1158-datove-zdroje-vav.html

Otevřená data – hlavní principy

- Strojová čitelnost (požadavky na formát)
- Nevylučující a jasně daná licence
- Katalogizace

První kroky TA ČR v oblasti otevřených dat

- Příprava publikování prvních datových sad: nejdůležitější: **stav projektů TA ČR** (napříč veřejnými soutěžemi)
  - Otevřenost: stupeň 3
  - Katalogizace – “katalog as a service”
- Později přibudou podkladové datové sady k infografikám, data ze systému INKAviz

Vize TA ČR v oblasti otevřených dat

- Využívání otevřených dat z dalších zdrojů (např. ČSSZ, ČSÚ – doplňování mozaiky dat např. o krajích)
- Využívání otevřených dat ostatních poskytovatelů podpory ve VaVaI (“které instituce neúspěšně podávající projekty v loňském roce se hlásí nyní k nám do veřejné soutěže…?”)

Nutnost vývoje aplikací nad otevřenými daty a spolupráce s dalšími institucemi…

T A
Č R

Děkuji za pozornost

www.tacr.cz