

T A

Č R

Výzkum užitečný pro společnost

T A

Č R

Využití shlukování textových dat v evaluacích – Helpdesk programů

Konference České evaluační společnosti 2021

10. 6. 2021

Technologická agentura České republiky

Martin Víta

Evaluační oddělení

- **Evaluace programů podpory VaVaI jsou standardní součástí jejich životního cyklu**
- Kromě *tradičních* metod (dotazníkových šetření, řízených rozhovorů, ...) se v TA ČR stále více **uplatňují data miningové (text miningové, graph miningové) přístupy**

- Oficiální **komunikační kanál** pro řešení dotazů potenciálních uchazečů, řešitelů, ...
- Pro každý z programů je k dispozici rubrika “*program*” – *veřejná soutěž* a “*program*” – *realizace*, jiné obsahové vymezení není

- Cílem analýzy Helpdesk je **identifikovat potenciálně problematické (příp. nedostatečně popsané/vysvětlené) části životního cyklu projektů**
- Předpoklad: **na problematické aspekty projektů se uchazeči/řešitelé ptají více než na ostatní**
- Často vyšší míra podrobnosti/konkrétnosti než při řízených rozhovorech

Základní idea analýzy Helpdesk I.

- Vytvořit shluky požadavků/dotazů z kolekce, která přináleží danému konkrétnímu programu
- Shluky jsou vytvářeny na základě “*sémantické podobnosti*”, tj. každý shluk obsahuje “*obsahově podobné dotazy*”
- **Shluky obsahující nejvíce dotazů/požadavků se odkazují na problémové okruhy**

Základní idea analýzy Helpdesk II.

Technologická agentura ČR > Helpdesk

Vyhledat položky

Informace pro tazatele

Helpdesk TA ČR slouží k zodpovídání dotazů týkajících se TA ČR, programů a veřejných soutěží TA ČR a dotazů směřujících na jiné poskytovatele využívajících informační systém ISTA (MV, MPO, MZa).

Pokud jste hodnotitel a máte dotaz/ požadavek týkající se hodnocení projektu, zašlete Váš dotaz e-mailem na adresu hodnotitele@taocr.cz.

Založit nový dotaz
Odešle Váš problém na technickou podporu

Zobrazit existující dotazy
Zobrazit dříve odeslané dotazy

Často kladené dotazy

Nejlepší články databáze znalostí:

What options of co-financing the project do a research organisation have? Can the missing resources on the research organisation side be co-financed by a project partner who is an enterprise? 7,166 ★★★★★ (32)

Kategorie: FINANCE

Yes, they can. It is up to the applicant and project partner(s) how they share the project costs. It is crucial that the aid intensity for the applicant/project partner or for the project cannot be ex...

Prohlášení k některým ustanovením všeobecných podmínek 5,204 ★★★★★ (37)

Kategorie: Realizace projektů

Prohlášení k některým ustanovením všeobecných podmínek je k dispozici zde.

Jaké dokumenty musím doložit před podpisem Smlouvy o poskytnutí podpory? 3,700 ★★★★★ (18)

Kategorie: Dokládání skutečnosti před podpisem Smlouvy o poskytnutí podpory

Výpis z Rejstříku trestů - zaslat datovou schránkou vyplněný dokument. Žádost o výpis z rejstříku trestů fyzická osoba , který je ke stažení na webových stránkách poskytovatele, popř. dodat dat...

Helpdesk TA ČR



T A
Č R

Program **Théta**

První shluk (interní číslo 3)

Největší shluk obsahující právě 24 dotazů/požadavků se týkal povinných či nepovinných příloh projektové žádosti a příloh dokládajících jednotlivé výsledky projektu, resp. jejich povinné formy odevzdání. Až na jednu výjimku byly všechny z dotazů zařazeny do kategorie THÉTA – veřejná soutěž.

Titulky jednotlivých dotazů / požadavků	Příklady dotazů
Nepovinná příloha projektové žádosti, povinná příloha – způsobilost, přílohy k žádosti o podporu, PP2 – povinná příloha k výsledku O, 3. výzva Théta, povinné přílohy PP2, Certifikovaná metodika, Povinné přílohy projektu, Výstup typu S – spec. veřejná databáze, Příloha, výsledek projektu, Navazující dotaz k ID:1412, Přílohy dle typu výstupu/výsledku, Výstup projektu – Certifikovaná metodika, Výstupy, Odevzdání certifikované metodiky, Povinné přílohy, povinná příloha, Popis výsledků Jimp v rámci typu O, Dotaz ke smlouvě o spolupráci a metodika, Výsledek, Smlouva o využití výsledků projektu, Přílohy podprogramu 3.	<p>Titulek: <i>přílohy k žádosti o podporu</i></p> <p>Text: Dobrý den, měla bych na Vás dotaz, zda je potřeba k žádosti předložit další přílohy k projektu kromě čestného prohlášení a Průzkumů trhu v případě, že výsledkem projektů bude prototyp? Je třeba doložit podrobný rozpočet a harmonogram k projektu? Předem děkuji za informaci a přeji hezký den ***</p> <p>Titulek: <i>Příloha</i></p> <p>Text: Krásný den, chtěla jsem se dotázat na povinné přílohy. U podprogramu č. 3 jsou povinné přílohy pouze pokud jsou výsledkem NmetS nebo Patent. Dále všichni účastníci projektu musí zaslat vyplněné čestné prohlášení uchazeče přes datovou schránku (příloha č. 1) Smlouva o účasti na řešení projektu mezi uchazeči není při podání návrhu projektu povinná, je jí potřeba nejpозději při uzavření smlouvy o poskytnutí podpory.</p> <p>Dotaz: existují ještě nějaké povinné či doporučené přílohy, které by vyžadovaly podpis statutárů uchazečů?</p> <p>...pověď a přeji krásný den.</p>

Evaluační zpráva

- **Obecná formulace úlohy:** je dána **kolekce entit reprezentovaných texty, cílem je vytvořit jejich shluky** na základě podobnosti dvojic entit
- **Dotaz/požadavek je reprezentovaný textem** “titulek + tělo dotazu/požadavku”

Posloupnost kroků procesu



- **Preprocessing textů** – zahrnující např.:
 - lemmatizaci – převody na základní tvar (“přehledu uznatelných nákladů” -> “přehled uznatelný náklad”,
 - odstranění stop slov (spojky, slova jako “projekt”),
 - odstranění bílých znaků, ...
- Texty reprezentujeme jako vektory (TF-IDF váhovaná DTM)

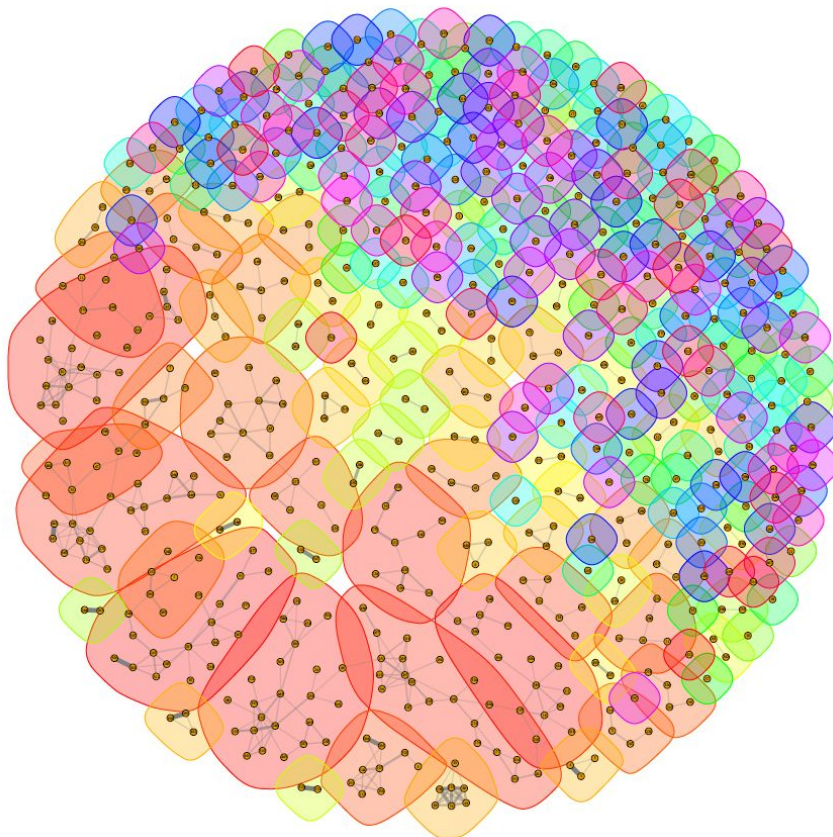
- Díky vektorové reprezentaci lze počítat **podobnost textů jako vzdálenost (vhodně zvolených) vektorových reprezentací**
- Zvolena tradiční **kosinová podobnost** (vzdálenost)
- Pro každou dvojici dotazů/požadavků tedy získáme jejich “vzdálenost” / podobnost (*můžeme získat matici vzdáleností, resp. graf – uzly: dotazy, hrany mezi nimi pak mají váhy odpovídající vzdálenostem/podobnosti*)

- Máme-li k dispozici vzdálenosti/podobnosti libovolných dvojic dotazů, lze **aplikovat clusterovací algoritmus (nabízí se řada možností)**
- Pro účely této analýzy byl zvolen algoritmus (WalkTrap) původně navržený pro identifikaci komunit na sociálních sítích

T A
Č R

Globální pohled na clusterování

Vizualizace grafu
s vyznačenými
clustery
(komunitami)
*Body (uzly) jsou
dotazy/požadavky*



T A
Č R

Získané clustery a jejich interpretace

Ukázky jednotlivých clusterů – popisy, interpretace
(separátní dokument)

Další příklady využití clusterové analýzy

- Analýza programů, resp. podprogramů – jaké tématicky podobné shluky projektů lze v programu identifikovat?
- ...

Děkuji za pozornost!

Dotazy, náměty, připomínky, prosím, zasílejte na
vita@tacr.cz