

Vedoucí týmu: prof. Petr Berka

Označení týmu: AI@VSE

1. Obsah výzkumu – state-of-the art

a. Data Mining a Knowledge Discovery: výzkum v této oblasti se zaměřuje na

- vývoj nových nástrojů pro práci s doménovou znalostí pro data mining,
- studium vlastností observačních kalkulů a jejich aplikace,
- vývoj nových nástrojů pro dolování z grafových dat (RDF)
- vývoj nových nástrojů pro automatizaci dolování z dat,
- výzkum v oblasti strojového učení, algoritmů a v oblasti srozumitelnosti výsledků data miningu,
- propojování data miningových systémů s business intelligence
- zpřístupnění „klasických“ data miningových nástrojů prostřednictvím webových rozhraní a sdílení výsledků data miningu přes web,
- extrakci strukturovaných dat z textů a analýzu sentimentu,
- aplikace data miningu mj. v oblasti byznysu, veřejné správy, zdravotnictví, a bezpečnosti informačních systémů.

b. Intelligent Information Systems: výzkum v této oblasti se zaměřuje na

- návrh a užívání systémů založených na logice, pravidlech a expertních systémů založených na příkladech a rozvoj bází doménových znalostí a bází znalostí pro specifické úlohy,
- modelování nejistoty v informačních systémech (fuzzy systémy, possibilistické modely a teorie informace),
- studium vztahů a rozdílů mezi lidskou a umělou inteligencí, simulační ověřování obecné umělé inteligence.

c. Semantic Web a Ontological Engineering: výzkum v této oblasti se zaměřuje na

- modelování propojených dat (linked data), jejich shromažďování, publikování a využívání zejména v oblastech

- veřejné správy, elektronického obchodu, akademické sféry, biomedicíny a encyklopedických znalostí (Wikipedia apod.),
 - návrh a správu ontologií a datových slovníků,
 - analýzu, evaluaci a mapování ontologií a datových slovníků,
 - vizualizaci ontologií a propojených dat,
 - dolování z propojených dat.
2. Klíčoví výzkumníci: prof. Petr Berka, doc. Tomáš Kliegr, prof. Jan Rauch, prof. Vojtěch Svátek, doc. Ondřej Zamazal
3. Klíčové metody a technologie:
- a. Metody
 - Data science
 - Strojové učení
 - Automatické vyvozování
 - Zpracování řeči a přirozeného jazyka
 - b. Technologie a aplikační oblasti
 - Bezpečnost a obrana
 - Internetové technologie
 - Asistivní a zdrav. technika (včetně oblasti biomedicíny)
 - Veřejná správa
4. Top 3 výsledky (systémy):
- LISp-Miner – data miningový systém pro výzkum a aplikace metody GUHA; zahrnuje procedury pro hledání bohatě strukturovaných asociačních pravidel, akčních pravidel a podmínek, za kterých kontingenční tabulky nebo histogramy splňují zadaná kritéria zajímavosti (<https://lispminer.vse.cz/>),
 - EasyMiner – open source webová aplikace pro učení asociačním pravidlům, klasifikaci s využitím pravidel a detekci anomálií (<http://easyminer.eu>),
 - EntityClassifier - nástroj pro vyhledávání entit v textu s využitím Linked Hypernym Dataset (<http://entityclassifier.eu/>).
5. Top 5 projektů (granty)
- GA CR 18-23964S „Fokusevaná kategorizační síla webových ontologií“, 2018-2020,

- EU H2020 OpenBudgets.eu, „Financial Transparency Platform for the Public Sector“, 2015-2017,
 - EU FP7 LinkedTV „Television Linked To The Web“, 2011-2015,
 - EU FP7 LOD2 „Creating Knowledge out of Interlinked Data“, 2010-2014,
 - GA CR 201/08/0802 “Application of Knowledge Engineering Methods in Knowledge Discovery from Databases”, 2008-2012.
6. Významný aplikační výstup či výsledek spolupráce s praxí nebo v praxi nasazen
- Použití systému LISp-Miner
 - i. ve společnosti HomeCredit pro pilotní analýzu spotřebitelských půjček
 - ii. ve firmě OptiSolutions pro vybrané analýzy reálných dat. Na základě zkušeností se systémem LISp-Miner byl do systému Lotylda vyvíjeného firmou OptiSolutions implementován modul OptiMiner zahrnující novou implementaci vybraných GUHA procedur pro data mining.
 - Public Contracts Ontology, v původní podobě nebo s úpravami se použila pro reálné publikování dat o veřejných zakázkách v Itálii a Španělsku
 - Česká větev strukturované webové databáze DBpedia (jako jednoho z nejvýznamnějších veřejných znalostních grafů) založené na Wikipedii