

# Znalostní a webové technologie co byste měli vědět

Doc. Ing. Vilém Sklenák, CSc.

Katedra informačního a znalostního inženýrství

7. 12. 2017

# Obsah

- 1 Kdo jsme
- 2 Znalostní a webové technologie
- 3 Co také děláme

# KIZI

- KIZI = **katedra informačního a znalostního inženýrství**
- jedna ze tří kateder FIS zaštiťující studijní program Aplikovaná informatika – tj. společně s:
  - katedrou informačních technologií (KIT)
  - a katedrou systémové analýzy (KSA).

# KIZI a bakalářský stupeň studia

- spolugarance oboru **Aplikovaná informatika** – již od jeho samotného počátku,
- garance oboru **Informační média a služby** – „mladý“ obor, poprvé od šk. roku 2016/2017.

# KIZI a bakalářský stupeň studia oboru AI

Povinné předměty:

- 4IZ110 – **Informační a komunikační technologie,**
- 4IZ210 – **Zpracování informací a znalostí.**

Volitelné předměty:

- 4IZ230 – **Umělá inteligence a reprezentace znalostí,**
- 4IZ260 – **Úvod do dobývání znalostí z databází,**
- 4IZ268 – **Webové technologie,**
- 4IZ278 – **Webové aplikace,**
- 4IZ238 – **XML – Teorie a praxe značkovacích jazyků.**

# KIZI a magisterský stupeň studia

Hlavní specializace:

- garance oboru **Znalostní a webové technologie**,
- podíl v oboru distančního studia **Podniková informatika**,
- podíl v anglickém oboru **Information Systems Management**.

Vedlejší specializace:

- garance specializace **4DI – Datové inženýrství** – není povolena pro studenty z programu AI,
- garance specializace **4IQ – Inteligentní systémy** – je povolena i pro studenty z programu AI.

# Obsah

- 1 Kdo jsme
- 2 Znalostní a webové technologie
- 3 Co také děláme

# Úvodem k ZWT

- V současné podobě se učí od šk. roku 2014/2015, navazuje však na předchozí obor **Znalostní technologie** a ještě dřívější **Informační a znalostní inženýrství**.
- Kořeny sahají až do 80. let 20. století, kdy se začala na VŠE rozvíjet tradice výzkumu v oblasti umělé inteligence či pokročilejších analýz dat.



# ZWT a jeho postavení v rámci AI

- ZWT je jeden ze 4 magisterských oborů – KIT (IST, KI), KSA (IM).
- Obor relativně malý – v průměru je počet přijatých každoročně kolem 20 studentů.
- Z nižšího počtu studentů plyne dobrý a úzký kontakt s katedrou, jehož výsledkem jsou pak třeba velmi kvalitní diplomové práce nebo zapojení studentů do výzkumných či vývojových projektů katedry.

# Základní profil

- Absolventi oboru získají pokročilé znalosti v oblasti všech fází vytěžování znalostí z různých datových zdrojů (firemních databází, informačních systémů, webů), tj. předzpracování dat různého charakteru, analýza, modelování, klasifikace, predikce a interpretace a následné prezentace získaných znalostí.
- Absolventi budou schopni navrhovat, přizpůsobovat, integrovat a realizovat systémy pro podporu rozhodování a integrovat je do informačních systémů.
- Dále se naučí vystavovat strukturovaná data z lokálních databází ve strojově čitelné podobě na web, a také vytvářet aplikace nad takto vystavenými daty.

# Komu je určeno

Obor je vhodný pro analyticky uvažující studenty, kteří by si rádi prohloubili své znalosti a dovednosti související se zpracováním potenciálně rozsáhlých dat (strukturovaných i nestrukturovaných).

# Možnosti uplatnění

- Absolventi oboru naleznou uplatnění ve větších i menších firmách jako specialisté v oblasti datových analýz (Business Intelligence), např. v získávání nových znalostí z dat specifických danému oboru, které pak vedení firmy použije při rozhodování a tvorbě nových strategií.
- Absolventi jsou důležitou součástí analytických týmů velkých firem, ale stejně tak se mohou uplatnit v institucích veřejné správy.
- Mohou nabízet softwarové řešení ušité na míru konkrétní firmě nebo jen služby v oblasti data miningu.
- Uplatnění na pozicích vyžadujících specializaci na data science.

# Povinné předměty

- 4IZ410 Teorie informace a inference
- 4IZ421 Metody a nástroje zpracování textových informací
- 4IZ430 Principy inteligentních systémů
- 4IZ450 Dobývání znalostí z databází
- 4IZ470 Dobývání znalostí z webu
- 4IZ440 Propojená data na webu
- 4IZ460 Pokročilé přístupy k DZD
- 4IZ501 Diplomový seminář

# Volitelné předměty

- 4IZ520 Aplikovaná lingvistika
- 4IZ524 Matematická informatika
- 4IZ525 Teorie kódování a šifrování
- 4IZ530 Logické programování a jeho aplikace
- 4IZ552 Elektronická sazba a publikování
- 4IZ560 Data mining – praktické aplikace
- 4IZ561 Umělé neuronové sítě
- 4IZ562 Řízení datové kvality
- 4IZ565 Data Science in Python and R<sup>1</sup>
- 4IZ5?? Webové služby/middleware<sup>2</sup>
- 4IZ617 Logika a sémantika
- 4IZ268 Webové technologie
- 4IZ278 Webové aplikace

---

<sup>1</sup>Nový předmět od LS šk. roku 2017/2018.

<sup>2</sup>Připravuje se od šk. roku 2018/2019.

## Volitelné předměty – pokračování

- 4EK415 Teorie rozhodování
- 4IT436 Business Intelligence
- 4MM405 Statistické metody pro analýzu dat z databází
- 4SA526 Nová média a sociální sítě
- 4ST210 Statistické výpočetní prostředí
- 4ST303 Analýza kategoriálních dat
- 4ST431 Časové řady
- 4ST512 Vícerozměrná statistika
- 5FI430 Znalosti a ontologické inženýrství

# Předměty vs. profil

- **Inteligentní systémy** – vyjma povinných předmětů předpokládá také 4IZ524, 4IZ617, 4IZ561, 4IZ525, 4EK415.
- **Znalostní a ontologické inženýrství** – k povinným předmětům ještě 4IT430, 4IZ617, 5FI430.
- **Zpracování textů a webu** – nad rámec povinných předmětů také 4IZ520, 4IZ538, 4SA220, 4IZ601, 4IZ602.
- **Data mining** – k povinným předmětům navíc 4IZ560, 4IZ562, 4ST303, 4ST431, 4ST512.



# Pochlubme se

DP Bohuslava Koukala „OLAP Recommender: Supporting Navigation in Data Cubes Using Association Rule Mining“, obhájená na oboru ZWT v r. 2017, postoupila do finále nejprestižnější česko-slovenské soutěže diplomových prací z oblasti informatiky IT SPY pořádané místními sekcemi asociace ACM (umístila se tedy mezi 9 nejlepšími pracemi z cca 1700 diplomek ze 16 zastoupených univerzit), viz <http://www.itspy.cz/en/projekty/>

Další success stories

# Obsah

- 1 Kdo jsme
- 2 Znalostní a webové technologie
- 3 Co také děláme

# Výzkum

V posledních 5 letech jsme se podíleli na 4 evropských projektech:

- EU Horizon 2020 Financial Transparency Platform for the Public Sector (OpenBudgets.eu), 2015–2017
- 7.RP EU LinkedTV – Television Linked to the Web, <http://linkedtv.eu>, 2011–2015
- 7.RP EU LOD2 – Creating Knowledge out of Interlinked Data, <http://lod2.eu>, 2011–2014
- 7.RP EU Multilingual-Web-LT, <http://www.multilingualweb.eu/>, 2012–2013

# Výzkumné skupiny

Na katedře jsou 4 neformální výzkumné skupiny:

- DMKD – Data Mining and Knowledge Discovery
- SWOE – Semantic Web and Ontological Engineering
- IIS – Intelligent Information Systems
- WELT – Web Engineering and Library Technology

# Semináře

„KEG semináře“ = 15 let trvající série nepravidelně konaných seminářů, které jsou veřejně přístupné:

- <http://keg.vse.cz>
- Program s tématy za celou historii konání

# Software

Výsledky z poslední doby:

- **EasyMiner** – webové rozhraní a REST API pro práci s asociačními pravidly,
- **InBeat** – software pro doporučení,
- **Targeted Hypernym Discovery** – práce s entitami extrahovanými z volného textu,
- **a ještě spousta dalších.**

# Spolupráce s praxí

V poslední době zájem o spolupráci s námi projevují mj. firmy MSD Innovation Center, BigML, Bisnode, Vendavo, Robert Bosch Research aj.

Děkujeme za pozornost.



## Úspěchy studentů Z(W)T

Diplomová práce Bohuslava Koukala „OLAP Recommender: Supporting Navigation in Data Cubes Using Association Rule Mining“, obhájená na oboru ZWT v r. 2017, postoupila do finále nejprestižnější česko-slovenské soutěže diplomových prací z oblasti informatiky **IT SPY** pořádané místními sekce asociace ACM (umístila se tedy mezi 9 nejlepšími pracemi z cca 1700 diplomek ze 16 univerzit), viz <http://www.itspy.cz/en/projekty/>. V rámci práce byl navržen a implementován softwarový systém **OLAP Recommender**, který kombinuje principy dataminingu a OLAP. Diplomant také systém s pozitivním ohlasem prezentoval v rámci vystoupení na česko-slovenské konferenci Data a znalosti v Plzni (2017).

Diplomová práce Petra Dušáka „Fractal application in data compression“, obhájená na oboru ZT v r. 2015, vznikla ve spolupráci s **Evropskou kosmickou agenturou**. Diplomant v jejím rámci vyvinul modifikaci patentovaného algoritmu pro kompresi časových řad, a ověřil její pozitivní vliv na míru komprese v rozsáhlém testu na datech z kosmické sondy Rosetta.

Diplomová práce Tomáše Hanzala „Modeling Events on the Semantic Web“, obhájená na oboru ZT v r. 2015, se stala základem dlouhého příspěvku na světové konferenci **Formal Ontology in Information Systems** konané v Annecy, Francie (2016).

Diplomová práce Petra Hazuzy „Ontologie přístupnosti budov“, obhájená na oboru ZT v r. 2015, vznikla ve spolupráci s **Kontem Bariéry**, kde její autor (vozíčkář) pracuje. V jejím rámci byl navržen rozsáhlý znalostní model umožňující strukturovaně popisovat míru přístupnosti budov a jejich částí pro tělesně postižené. Znalostní model se využívá v projektu *Mapy bez bariér*, viz <https://mapybezbarier.cz>.

Diplomová práce Šárky Turečkové „Využití propojených dat na webu ke tvorbě strategické znalostní hry“, obhájená na oboru ZT v r. 2015, byla formou příspěvku s dobrým ohlasem prezentována na česko-slovenské konferenci Data a znalosti (2015). Práce využívala **znalostní databázi DBpedia** jako vstupní data pro strategickou hru umožňující současně vzdělávání ve zvolené oblasti.

Dva studenti oboru ZWT, B. Koukal a L. Horáková, se během let 2015-2017 jako placení spolupracovníci zúčastnili řešení **projektu OpenBudgets.org** (<https://openbudgets.eu>) financovaného z prestižního programu **EU Horizont 2020**. Podíleli se na tvorbě datových modelů a analýze dat v oblasti veřejných rozpočtů. Některé výsledky byly použity i v jejich diplomových pracích.

Diplomová práce Marka Dudáše, obhájená na oboru ZT v r. 2013, se postupně stala základem dvou prestižních mezinárodních publikací (dokončených už v rámci doktorského studia): dlouhého příspěvku v hlavním programu **evropské konference o znalostním inženýrství EKAW** (2014, přijímalo se cca 20% z nabídnutých) a článku v **americkém impaktovaném časopise Knowledge Engineering Review** (2017).

V mezinárodní soutěži systémů pro mapování ontologií, OAEI, konané při světové konferenci o sémantickém webu, ISWC, se od r. 2006 využívá **benchmarková kolekce ontologií OntoFarm**. Její základ vznikl jako soubor semestrálních prací studentů (zejména) oboru ZT v předmětu 4IZ440.

Tři z absolventů oboru ZT získali krátce po obhajobě (už jako doktorandi) **Cenu J. Hlávky** pro mladé vědce: Ondřej Zamazal (2006), Tomáš Kliegr (2009) a Marek Dudáš (2015). Všichni v současnosti dále působí na katedře.