

Digitální obrat v humanitních vědách

Michal Lorenz

Katedra informačních studií a knihovnictví

LINDAT/CLARIAH-CZ

Co jsou digitální humanitní vědy?

obor zkoumající vztahy mezi:

- praxí humanitních oborů
- *(práce vědců, informační potřeby a požadavky)*
- digitálními zdroji
- *(elektronické texty, digitální fondy, nástroje)*
- metodami
- *(nomotetické, počítačové metody)*
- technologiemi použitelnými pro zodpovězení výzkumných otázek

Digital Humanities (DH) - úvod

- Humanitní obory – studují lidskou kulturu a její doklady (produkty, které za sebou lidstvo zanechává: artefakty, obrazy, texty, digitální objekty)
- „Digitální humanitní obory“ – centrum UCLA zahrnuje akademické zaměstnance z širokého spektra disciplín: od humanitních věd přes vědy sociální, přírodní a technické
- rostoucí množství digitálních reprezentací materiálních objektů – rozšiřování žánrů dostupné evidence
- v humanitních oborech zažívají explozi dat jako jiné obory, jen takto nebývá pojmenována
- data – entity používané jako evidence fenoménu pro účely výzkumu a vědy
- původně ideografické metody vysvětlení se s rostoucím množstvím dat posunují k vysvětlením založeným na nomotetických metodách

Digitální (výpočetní) obrat

- digitální technologie mění výzkumný proces – ovlivňují epistemologii i ontologii
- kontinuální tok reality - diskrétní kódování – algoritmická manipulace dat: substraktivní metody způsob poznávání reality
- reprezentace a mediace digitálních forem obsahu – nutná digitální či výpočetní technika
- výpočetní technika podmínkou promýšlení otázek humanitních studií
- software umožňuje nový způsob kritického čtení a psaní (kódy, vizualizace dat, vzorů a vyprávění), nové metody pedagogiky
- výpočetní obrat usnadňuje disciplinární hybridy, které vedou ke vzniku post-disciplinární univerzity

Transformace dat v Digital Humanities

- **11. fyzické x digitální objekty:** vědci zkoumající materiály – problém s knihovnami a archivy – přístup k vzácným objektům x vědci zkoumající obsah – lepší digitální reprezentace: možno srovnání, vyhledávání, tištění, anotování
- **2. surogáty x úplný obsah:** bibliografické záznamy, miniatury obrázků, audio a video klipy a další reprezentace objektů – pomáhají rozšiřovat, popisovat, lokalizovat, vyhledávat a autentizovat digitální obsah, nemění zásadně vědecké praktiky x digitalizovaný obsah podle materiálu může být vytěžován, kombinován, modelován, využíván jako zdroj dat
- **3. statické x prohledavatelné reprezentace:** využití digitálního obsahu otázkou stupně – od statických obrázků po sémanticky obohacené objekty, písmena a text prohledavatelné díky OCR (původ v telegrafickém systému 20. stol., problém s historickými knihami a záznamy – font, pozadí)
- **4. prohledavatelné řetězce x obohacený obsah:** nové výzkumné metody a zkoumání otevřeny díky strukturovaným reprezentacím digitálních dokumentů, standart TEI (Text Encoding Initiative) od 80. let rozšířen na nástroje pro Latinu, Řečtinu, hudbu atd. Sémantické informace o slovech, frázích, nadpisech, kapitolách, stránkách, mezerách, osobních jménech, místech, poznámkách, tagy a zakódování obrazů, audia i videa, rentgenové techniky pro zobrazení palimpsestů
- př. Paris – město, jméno řeckého boha, forma sádry
- Brown – barva, jméno
- BLAST – nástroj pro prohledávání sekvencí DNA používán i v humanitních studiích k prohledávání řetězců digitálních objektů

Vrstvy sémantického webu

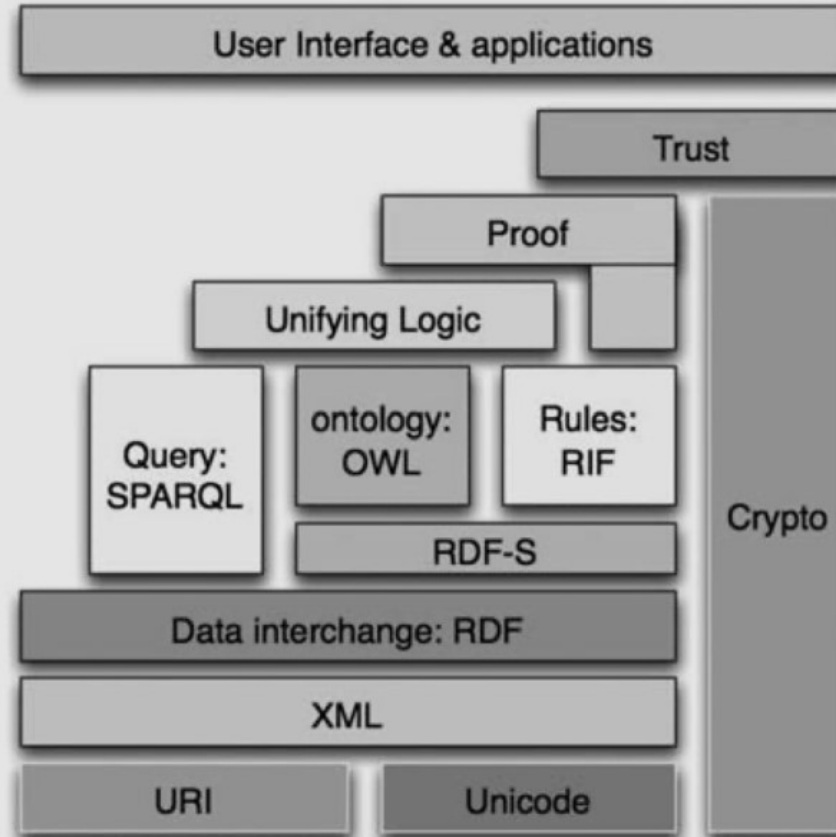


FIGURE 24.1 The layered view of the semantic web. (From A framework for Web Science, by T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt, D.J. Weitzner, Found. Trends Web Sci. 2006, 1 (1), 1–134.)

Metody v DH

- **Linked data a RDF**
- **TEI** (Text Encoding Initiative)
- př. Paris – město, jméno řeckého boha, forma sádry
- Brown – barva, jméno
- **BLAST** – nástroj pro prohledávání sekvencí DNA používán i v humanitních studiích k prohledávání řetězců digitálních objektů
- **LIDAR** - https://youtu.be/eBUCGxZq_xg

Témata

- Digitalizovaná bibliografická data – primární a sekundární zdroje, ale i generační skladba spisovatelské obce, ohlas literárních děl..
- Korpusy – studium společenských jevů indikovaných jazykem (př. Rapcor) nebo sémantika barevného spaktra u spisovatelů žánrů – autorská práce s barvami, srovnání s kognitivními významy ve velkých korpusech..
- Sémantický prostor biografií – zkušenosti mužských a ženských vypravěčů
- Sémantické sítě – matice vazeb mezi slovy na základě spoluvýskytu: morální termíny v korpusech antických textů – např. kulturní evoluce náboženství
- Modelování témat – automatické zjišťování námětů témat
- Geografické informační systémy a sociální sítě
- Digitální etnografie
- 3D vizualizace

LINDAT/CLARIAH-CZ

- Digitální výzkumná infrastruktura pro humanitní, jazyková a umělecká data

<https://lindat.cz/>

- Katalog platforem a digitálních sbírek na FF MU:

<https://it.muni.cz/phil/lindat-clariah/katalog>

Původ DH

- 1949 – textová analýza: jezuita Roberto Busa se spojil s IBM, aby digitalizoval kompletní dílo sv. Tomáše Akvinského, později i Bibli, Kumránské spisy i Korán
- 90. léta – zájem o hypertext, digitální repozitáře, multimediální sbírky
- 21. století – sociální sítě, crowdsourcing, digitalizované kulturní dědictví, rozhraní pro archivy a sbírky využívající propojená data (linked) a vizualizaci
- DH často spojeno s katedrami anglistiky – historické spojení s textem, výpočty, kompozicí, editorské procesy, hypertext a kulturní studia

Scholarly primitives

- Pojem zavedl J. Unsworth v roce 2000
- fungují jako axiomy, na nichž stavěny vědecké argumenty, tvrzení a interpretace
- seznam základních funkcí vykonávaných vědci během vědecké činnosti, opakujících se napříč vědeckými obory
- důležité pro design digitálních nástrojů a elektronických rozhraní

Table 2: The five core scholarly activities and their primitives

1. Searching

1.1 Direct searching

1.2 Chaining

1.3 Browsing

1.4 Probing

1.5 Accessing

2. Collecting

2.1 Gathering

2.2 Organizing

3. Reading

3.1 Scanning

3.2 Assessing

3.3 Rereading

4. Writing

4.1 Assembling

4.2 Co-authoring

4.3 Disseminating

5. Collaborating

5.1 Coordinating

5.2 Networking

5.3 Consulting

6. Cross-cutting Primitives

6.1 Monitoring

6.2 Notetaking

6.3 Translating

6.4 Data Practices