



Digitální obrat v českých humanitních a sociálních vědách

Radim Hladík
a kolektiv

Digitální obrat v českých humanitních a sociálních vědách

Radim Hladík a kolektiv

Recenzovali:

Mgr. Marek Debnár, Ph.D.

Mgr. Martin Charvát, Ph.D.



**Financováno
Evropskou unií**
NextGenerationEU



**Národní
plán
obnovy**

**MS
MT**
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Publikace byla vydána za podpory Ministerstva školství, mládeže
a tělovýchovy a Národního plánu obnovy v rámci projektu
Transformace pro VŠ na UK (reg. č. NPO_UK_MSMT-16602/2022).

Vydala Univerzita Karlova
Nakladatelství Karolinum
Praha 2022
Redakce Václav Hozman
Grafická úprava Jan Šerých
Sazba DTP Nakladatelství Karolinum
Vydání první

© Univerzita Karlova, 2022

© Radim Hladík a kolektiv, 2022

ISBN 978-80-246-5193-4

ISBN 978-80-246-5393-8 (online : pdf)



Univerzita Karlova
Nakladatelství Karolinum

www.karolinum.cz
ebooks@karolinum.cz

OBSAH

Předmluva <i>Radim Hladík</i>	7
Zdroje a nástroje	
Možnosti využití bibliografických datasetů jako zdrojů pro kvantitativní výzkum v humanitních vědách: případ České literární bibliografie <i>Vojtěch Malínek</i>	19
Moderní popis staročeské morfologie <i>Boris Lehečka, Pavlína Synková, Ondřej Svoboda, Jiří Pergler, Jana Zdeňková</i>	45
Přínos jazykových korpusů pro studium charakteristik neformální mluvené komunikace <i>Zuzana Laubeová, Marie Kopřivová, Michal Křen, Petra Poukarová</i>	67
Strojové čtení	
Počítačová analýza vyprávění <i>Martin Hájek</i>	89
Řízená literárněvědná interpretace fikční sémantiky: přínos digitálních metod literární vědě <i>Richard Změlík</i>	109
Kookurenční sítě mezi historickým vyprávěním a kulturněevolučním vysvětlením <i>Vojtěch Kaše</i>	133
Modelování témat v české sociologii: typy autorství a citační ohlas v odborných textech <i>Radim Hladík</i>	159

Dějiny v datech

Vědecké blogování a geografie holokaustu: Document Blog EHRI
Michal Frankl 193

Česká šlechta a její sociální a prostorové vazby:
využití geoinformačních a síťových dat v medievistice
Jan Škvrňák, Adam Mertel 213

Trojrozměrné digitální rekonstrukční modely v archeologii
Jiří Unger, Petr Květina, Jan Mařík 231

Krajina očima archeologie s digitální optikou
Martin Kuna, David Novák 259

Digitální současnost

Zdraví a zdravotnictví v digitální éře
Dino Numerato 297

Polarizace v internetové diskusi: analýza sociálních sítí
s negativními vazbami
Matouš Pilnáček, Jaromír Mazák, Tomáš Diviák 315

Digitální muzea aneb muzea beze zdí
Nina Wančová 343

Výuka digitálních humanitních věd na českých veřejných
vysokých školách podle latentní sémantické analýzy
Silvie Cinková, Jan Škvrňák, Michael Škvrňák 367

Seznam autorů 409

PŘEDMLUVA

Radim Hladík

Ohledně vymezení digitálních humanitních věd (*digital humanities*, DH) a komputačních sociálních věd (*computational social sciences*, CSS) panuje tolik názorů, že snahy o jejich definici tvoří samostatný žánr. Vzpírají se jasnému vymezení, neboť se samy průběžně vyvíjejí. Jejich dynamika, která se snaží držet krok s rychlými pokroky ve vývoji nových algoritmů, nepřeje tomu, aby digitální přístupy vnesly do humanitního a sociálněvědního výzkumu pevně nastolené výzkumné programy, s nimiž by institucionalizované obory mohly standardně spolupracovat.

Při úvahách o tom, jak digitální humanitní a sociální vědy představit českým výzkumníkům a studentům, by příklon k té či oné definici mohl napáchat více škody než užitku. Ještě méně užitečné by bylo pokoušet se přidat k již existujícímu dlouhému seznamu definic vlastní. Tím nemá být řečeno, že úctyhodná intelektuální práce, jež byla dosud v teorii digitálních humanitních věd vynaložena, přichází vniveč. Pro budoucí rozvoj tohoto oboru zůstávají takové kritické diskuse naprosto nezbytné. Tato kniha se však k představení digitálních humanitních věd vydává oklikou. Ve svém zaměření na demonstrativní případové studie vychází z úsloví, že důkaz pudinku je v jídle.

Autorky a autoři, kteří odpověděli na výzvu představit principy digitálního bádání širší akademické komunitě, přijali netriviální úkol ukázat čtenářům, jak oni samotní nebo jiní badatelé postupují při zodpovídání výzkumných otázek spojených s digitálními přístupy, aniž by použité metody popisovali technickým žargonem a for-

mulemi. Pokud jsou metody, jakkoliv by mohly být jejich výsledky zajímavé, prezentovány jako černá skříňka, nepůsobí přesvědčivě. Na druhou stranu, přísně zdokumentovaná metodologie může na netrévaného čtenáře působit esotericky a nepřístupně. Proto zde dáváme nahlédnout do principů digitálního bádání tak, jako bychom servírovali pudink bez receptu, ale s otevřenými dveřmi do kuchyně.

Navzdory svému praktickému zaměření tato kniha nemá sloužit jako autoritativní učebnice nebo příručka k osvojení metod. Zvolený hybridní žánr inklinuje k didakticky pojatým původním případovým i přehledovým studiím, jejichž soubor cílí na konstrukci od technických detailů abstrahovaného intelektuálního modelu toho, co znamená provádět digitální výzkum na empirickém materiálu humanitní a sociálněvědní provenience. Pokud se autorskému kolektivu podařilo zadaný úkol splnit, čtenáři by si měli z četby vedle dílčích poznatků o demonstrativních výzkumných problémech odnášet především pochopení toho, jak digitální humanitní vědy uvažují o svých výzkumných postupech, jak formulují své otázky a problémy a jakými transformacemi procházejí jejich předměty bádání.

Induktivně vybudovaný intelektuální model digitálního výzkumu, který zde nahrazuje teoretické definice, nevyžaduje, aby digitální humanitní a sociální vědy vytvářely vlastní oborové teritorium. Jednotlivé studie fungují spíše jako špendlíky zapíchnuté do již známých map, kde označují místa specifického badatelského zájmu. Samotná šíře tematického spektra pokrytého v této knize pak dokládá, že digitální postupy nejsou vlastní žádné specializaci nebo tématu.

Variabilita uplatnění digitálních témat a přístupů se odráží i v členění jednotlivých kapitol. Bylo možné se v něm opřít i o tradiční oborové nomenklatury a seskupit příspěvky podle jejich mateřských disciplín. Zastoupeny jsou z této perspektivy lingvistika, lexikografie, sociologie, archeologie, historiografie, mediální studia, religionistika, literární věda i muzeologie. Konečná klasifikace se však opírá o jiná pravidla, která zdůrazňují interdisciplinární spojení různých technik a problémů.

Důvod, proč digitální humanitní vědy a počítačnické sociální vědy často vystupují nebo jsou vnímány jako svébytné specializace, může spočívat právě ve vzájemných vazbách mezi různými technikami a datovými modely než v existenci autonomních výzkumných

programů. Digitální sociolog i literární vědec se mohou od sebe inspirovat a sdílet zájem o analýzu textů navzdory tomu, že každý z nich se zabývá jiným empirickým materiálem. Digitální historik může využít nástrojů vyvíjených počítačovými lingvisty. Religionista i archeolog mohou k různým účelům využívat stejný geoinformační systém. Struktura knihy se tedy snaží sledovat tyto na první pohled skryté souvislosti, které mnohdy vycházejí spíše z povahy analyzovaných dat než z oborové příslušnosti.

První sekce, *Ždroje a nástroje*, obsahuje příspěvky, jež spojuje zaměření na tvorbu dat a nástrojů pro jejich zpřístupňování. Tyto datové platformy např. v podobě databází, korpusů či slovníků jsou označovány jako výzkumné infrastruktury, neboť podobně jako třeba laboratorní přístroje fungují jako základna pro další výzkum. Hranice mezi infrastrukturou pro výzkum a výzkumem samotným však nemusí být ostrá, jelikož vznik infrastruktur se často odvíjí od badatelské práce.

Kapitola o moderním popisu staročeské morfologie čtenáře upozorňuje na množství problémů, se kterými se musí potýkat jazykovědci při vymýšlení optimálních datových modelů pro starou češtinu. Zatímco obvykle uživatelé vidí pouze výsledné aplikace, kolektiv autorů vedený Borisem Lehečkou pro ně odkrývá množství dílčích kroků a designových rozhodnutí, které předchází finální prezentaci a poskytování dat.

Další kapitoly kromě způsobů budování infrastruktur demonstrují také možnosti jejich výzkumného využití. Vojtěch Malínek popisuje přerod souboru bibliografických záznamů o české literatuře do datové podoby a testuje nové možnosti, které digitalizovaná bibliografická data nabízejí. V elektronické podobě se využití bibliografické databáze nemusí omezovat pouze na vyhledávání primární a sekundární literatury, ale má bezprostřední význam i pro kvantitativně podložené studie o generační skladbě české spisovatelské obce nebo o ohlasu literárních děl.

Podobné propojení infrastruktur s výzkumem demonstruje Zuzana Laubeová se svými spoluautory, kteří vyzdvihují přínos digitálních korpusů pro lingvistické bádání. Počítačově zpracované korpusy lingvistům umožňují zkoumat řadu jazykových jevů, a to na širokém spektru uživatelů jazyka a v různých kontextech. Příkladem jazykového vývoje mohou být změny ve slovnědruhovém klasifi-

kaci v mluvené češtině. Korpusová lingvistika otevírá prostor i pro studie, v nichž jazykové jevy mohou indikovat společenské procesy.

Analýza textových dokumentů vystupuje jako součást výzkumu v mnoha oborech. Dokument z hlediska digitálních humanitních a sociálních věd může nabírat různou podobu. Badatelé mohou analyzovat otevřené otázky v sociologických nebo psychologických dotaznících stejně jako středověké rukopisy nebo literární texty od poezie po romány. Otázce využití počítačů v analýze dokumentů se věnuje oddíl *Strojové čtení*.

Martin Hájek pomocí počítačových skriptů mapuje sémantický prostor biografických příběhů zaznamenaných v sociologických rozhovorech. Jeho kapitola se věnuje různým postupům pro rozbor vyprávění a formálnímu zachycení této všudypřítomné diskursivní formy. Ukázková studie odhaluje odlišně vnímané životní zkušenosti mužských a ženských vypravěčů, kteří vzpomínají na dobu státního socialismu a jejichž výpovědi jsou uloženy v orálněhistorických archívech.

Uplatněním digitálních metod v literární vědě se věnuje Richard Změlík. Právě z oblasti literární vědy se rekrutují již klasické výzkumy např. Franca Morettiho, jehož koncept vzdáleného či distančního čtení silně ovlivnil veřejnou představu o digitálních humanitních vědách, ale také vyvolal řadu polemik o schopnosti počítačů podílet se na interpretaci literárních textů. Čtenářům příspěvek nabízí praktickou ukázkou toho, jak lze využít metody korpusové lingvistiky k daty řízenému výkladu sémantiky barevného spektra, s nímž pracují spisovatelé při vytváření fikčních světů. Práce s vícero korpusy a textovými metadaty (tedy daty, kterými popisujeme data) umožňuje také porovnávat autorsky podmíněnou práci s barvami nejen mezi jednotlivými spisovateli, ale i v kontrastu s kognitivními významy barev, které jsou zachyceny ve velkých jazykových korpusech.

Sémantické sítě jsou matice, které kvantifikují vazbu mezi slovy na základě jejich spoluvýskytu v definovaném textovém úseku. Za pomoci těchto sítí extrahovaných z korpusu antických textů se Vojtěch Kaše snaží sledovat frekvence a významové okolí morálních termínů, aby posoudil vzájemně konkurenční hypotézy, jež o kulturní evoluci náboženství postulují religionisté. Výsledky naznačují, že představy o rozvoji moralizujících náboženství v souvislosti s nástupem křesťanství nemají velkou empirickou oporu v dochovaných

textech, které místo toho dokládají silné propojení morálky a náboženství již ve starších obdobích.

Editor monografie se věnuje dalšímu typu dokumentů, kterým jsou odborné vědecké texty. Jeho kapitola hledá vztah mezi genderem autorů, počtem autorů i citovaností sociologických článků na straně jedné a tématem článků na straně druhé. Analýza se opírá o modelování témat, což je technika, kterou digitální humanitní vědy často používají pro automatické zjišťování námětu textů. Tematický model odhaluje, že vnitřní rozrůzněnost sociologie není náhodná, ale odvíjí se od charakteristik autorů a jejich způsobu práce. Nalezená asociace mezi citovaností a tématy článků také upozorňuje na nedostatečnou vypovídací hodnotu agregátních časopiseckých statistik, jež jsou v hodnocení vědy často užívanými metrikami.

Dějiny v datech je název oddílu zaměřeného na příspěvky využívající digitální metody pro studium historie. Michal Frankl uvažuje o rozšíření perspektivy, kterou historikům holokaustu nabídl blogová platforma EHRI Document Blog. Schopnost internetového média kombinovat různé typy dat, začleňovat do textu nejen obrázky, ale i interaktivní mapy a jiné vizualizace stimuluje nové formy historiografického bádání. Flexibilita blogu vede k intenzivnější komunikaci mezi historiky a poskytuje prostor pro prezentaci dílčích pokroků v bádání, o nichž lze dále diskutovat. Zpětná vazba vede k cizelování publikovaných studií předtím, než jsou závěry publikovány v tradičních časopisech. Provoz blogu však také přináší technické výzvy, jež vyžadují, aby badatelé refleктоvali povahu dat, se kterými pracují.

Přehled několika způsobů, jimiž mohou digitální metody přispět k medievistickému výzkumu, sepsali Jan Škvrňák a Adam Mertel. Historická data lze stejně jako ta současná podrobit zkoumání pomocí geografických informačních systémů nebo sociálních sítí. Propojení obou perspektiv pak dovoluje sledovat takové výzkumné otázky, jako zda geografická blízkost ovlivňovala sociální vazby středověké šlechty nebo jak pozice vládce v sociální síti aristokracie indikuje politický smír. Na jednu stranu mezi historickými a současnými daty neexistuje zásadní rozdíl z hlediska metod, kterými mohou být analyzována. Na straně druhé je při práci s historickými daty potřeba brát pečlivě v úvahu jejich fragmentárnost a zkrácení vyplývající z jejich nenáhodné dostupnosti a zachovalosti. Autoři

proto dávají prostor i úvahám o tom, jaké okolnosti musí mít badatelé na paměti, když historické prameny transformují do podoby digitálních historických dat.

Dvě kapitoly zaměřené na výzkum minulosti pocházejí z oboru archeologie. V tomto oboru se digitální obrat v rámci humanitních a sociálních věd projevuje možná nejmarkantněji. Přispívá k tomu skutečnost, že archeologové pracují v terénu, a tak nové technologie hojně uplatňují nejen při analýze, nýbrž i při sběru dat. Příklady těchto činností přináší Martin Kuna a David Novák, kteří čtenářům přibližují, jak digitální archeologie přetváří krajinu do počítačem zpracovatelného objektu. Základním datovým modelem pro digitalizaci krajiny je geografický informační systém, který eviduje prostorová data v různých vrstvách a rozměrech a v němž mohou být zachyceny výskyty archeologických lokalit a jakékoli další body zájmu. Prostorová data mohou být obohacena i leteckým laserovým skenováním terénu, díky němuž lze analýzou výškopisných modelů v krajině hledat dosud neznámé stopy minulosti. Archeologové tak mohou získávat poznatky o sídlištních strukturách, o prostorovém rozptýlení historických populací nebo o ekonomických systémech v různých obdobích.

Trojrozměrnými vizualizacemi archeologických dat se zabývá kolektiv autorů v čele s Jiřím Ungerem. Grafická znázornění archeologických objektů mohou sahat od drobných předmětů až po celá města a využití 3D modelů se v archeologii rychle rozšířilo. Pokročilé modely vyžadují práci vysoce odborných a interdisciplinárních týmů, které zahrnují i počítačové vědce a umělce. Ani technická sofistikovanost však nepřinesla, jak autoři upozorňují, kýžený objem nových poznatků, jež by bylo možné z vizualizací odvodit. Zcela nezastupitelné místo však počítačem vykreslené modely získaly pro komunikaci archeologického poznání širší veřejnosti. I přes své ukotvení v archeologii nese tato kapitola důležité poselství i pro další obory, neboť se nebrání epistemologickým úvahám o obecném významu vizualizací pro vědecký výzkum. Počítačová grafika dává humanitním a sociálním vědcům k dispozici mnoho nástrojů k prezentaci jejich dat a může posilovat snahy o popularizaci jejich vědecké práce. Zdánlivá samozřejmost vizualizací však skýtá i nástrahy v situacích, kdy data, nad nimiž vizualizace stojí, jednoznačně zdaleka nejsou. Proto se autoři věnují i otázce vizualizace výzkumné nejistoty.

V posledním oddílu jsou shromážděny příspěvky, jejichž náplň jde nad rámec reflexe digitálního obratu v odborných disciplínách. Reagují na aktuální svět, který sám v překotném tempu prochází vlastní digitalizací. Oddíl *Digitální současnost* otevírá kapitola Dina Numerata o zdraví a zdravotnictví v digitální éře. Podobně jako jiné oblasti života i péče o zdraví se více a více uchyluje k novým technologiím. Povinností sociologů je tyto proměny monitorovat a Dino Numerato předkládá výběrový, a přesto rozsáhlý výčet společenských pojmů, které pod vlivem digitalizace zdraví nabírají zcela nový obsah. Tradiční vztah mezi medicínskými experty a pacienty narušuje dostupnost informací na internetu, který pacienti navíc využívají ke sdílení svých osobních zkušeností. Informovanost, důvěra a kontrola mezi pacienty a experty hledají novou rovnováhu na pozadí digitalizace samotné medicíny, která díky velkým souborům dat, bioinformatice a strojovému učení testuje nové způsoby péče a léčby jednotlivců i celých populací.

Rozpad veřejnosti do polarizovaných názorových skupin patří k nejvýraznějším politickým fenoménům digitální současnosti: Podílí se na něm sociální síť ve významu platform, na nichž se ve virtuálním prostoru setkávají a sebe prezentují jednotlivci. Termín sociální síť však také označuje odborný koncept, s jehož pomocí sociální i humanitní vědci zkoumají vzájemné vazby různých společenských aktérů. Matouš Pilnáček a jeho kolegové aplikují analýzu sociálních sítí na výzkum politické polarizace tématu migrace v internetových diskusích mezi čtenáři zpravodajského portálu. Kapitola detailně vysvětluje, co obnáší analytický přístup založený na síťové perspektivě. Aby hrubá data z čtenářských komentářů převedli do relační formy, museli v nich výzkumníci indentifikovat vztahy pomocí hodnocení komentářů, která si diskutující mezi sebou udělují. Po automatizovaném sběru dat a jejich transformaci mohou již výzkumníci přikročit k využití grafových statistik, aby vyhodnotili míru názorové polarizace na diskusním fóru. Výsledky napovídají, že v diskusích o migraci neexistují dvě, nýbrž tři skupiny diskutujících, přičemž hlavním impulsem pro polarizaci debaty jsou komentáře postavené na aktivně protiimigrační rétorice.

Nina Wančová shrnuje poznatky o nových způsobech prezentování muzejních artefaktů v digitální podobě. Virtuální expozice neposkytují pouze další výstavní prostor, ale vedou k nekonvenčnímu

promyšlení funkcí a poslání muzejních sbírek. Digitalizované exponáty a jejich popisy v podobě metadat už nejsou pouze výsledkem badatelské a kurátorské práce, ale mohou se stát i infrastrukturami pro další výzkum, jak o tom primárně pojednávají úvodní příspěvky této knihy. Kapitola Niny Wančové mimo jiné podává zprávu o zkušenostech s vývojem a uplatněním českého softwarového nástroje pro digitální muzea INDIHU Exhibition. Při vývoji tohoto softwaru bylo nutné nalézt optimální řešení zohledňující jak kurátorské cíle, tak uživatelská očekávání. Výsledná asembláž multimediálních formátů nenahrazuje virtuálním návštěvníkům fyzickou muzejní expozici, ale poskytuje jim svébytný styl setkávání se se sbírkovými předměty.

Závěrečná studie Silvie Cinkové a dalších autorů završuje směřování knihy k orientační mapě českých digitálních humanitních a sociálních věd. Po metodologické stránce se příspěvek opírá o latentní sémantickou analýzu, která podobně jako modelování témat třídí dokumenty podle obsahové příbuznosti. Jednotlivé kroky této metody jsou detailně popsány a nastíněny jsou i cesty jejího rozvíjení dodatečným zpracováním dat pomocí shlukové a síťové analýzy. Náplní explorativního výzkumu jsou studijní katalogy českých vysokých škol, v nichž autoři hledají předměty, jejichž výuka může být relevantní pro digitální humanitní vědy. Ačkoliv studie metodologicky doplňuje spíše sekci *Strojové čtení*, její substantivní téma představuje navýsost aktuální přehled o přítomnosti výuky digitálních metod v českém terciárním vzdělávání. Závěry autorů jsou v zásadě optimistické – přestože předmětů, které by se explicitně věnovaly digitálním humanitním vědám, je poskrovnu, pro dnešní studenty již existuje poměrně široká nabídka kurzů digitálních dovedností, které lze v digitálním výzkumu uplatnit. Poslední kapitola tak přináší empirickou oporu pro leitmotiv celé knihy, kterým je přesvědčení, že bez ohledu na to, zda si digitální humanitní vědy a počítačové sociální vědy získají v Česku robustní institucionální zázemí, digitální obrat nevyhnutelně rezonuje napříč obory.

Zvolené osy – infrastruktury, texty, historická data a současné problémy –, kolem nichž se jednotlivé příspěvky soustředí, zachycují jen vybrané rozměry mnohorozměrného procesu, kterým je přejímání digitálních metod a jejich aplikování ve výzkumu v humanitních a sociálních vědách. Kromě již předestřené oborové příslušnosti

mohou být příspěvky propojeny i metodologickými kritérii. Čtenáři si mohou všimnout, že některé příspěvky spojuje zájem o analýzu sociálních sítí, jiné využívají různé varianty maticových operací, frekvenční statistiky, vizualizace nebo spoléhají na geoinformační systémy či přehledy literatury. Knihu lze tedy číst vícero způsoby a sestavit si vlastní řazení příspěvků.

I přes svůj široký záběr a tvárnost vlastní organizace zanechává kniha mnoho slepých míst. Určitě v ní schází příspěvky psychologické, ekonomické, antropologické, muzikologické, stylometrické a další typy příspěvků. Cílem knihy však není poskytnout vyčerpávající přehled všech metod a oborů, nýbrž pouze nastínit komplexní charakter digitálního obratu v českých humanitních a sociálních vědách. Pokud budeme vnímat digitální obrat ve výzkumu v celé jeho složitosti a mnohostrannosti, lépe pochopíme i potíže s jeho nestabilními teoretickými definicemi. Ponechme tentokrát velké teorie stranou a pokusme se přemýšlet nad tím, zda ve svém portfoliu výzkumných otázek nemáme takové, kde bychom pomoc digitálních metod uvítali. Přednosti i limity těchto metod se nejvýrazněji zhmotní právě při jejich kritickém použití. Je-li důkaz pudinku v jídle, pak vám při čtenbě přeji dobrou chuť.



ZDROJE A NÁSTROJE

MOŽNOSTI VYUŽITÍ BIBLIOGRAFICKÝCH DATASETŮ JAKO ZDROJŮ PRO KVANTITATIVNÍ VÝZKUM V HUMANITNÍCH VĚDÁCH: PŘÍPAD ČESKÉ LITERÁRNÍ BIBLIOGRAFIE

Vojtěch Malínek

Bibliografie fungují po dlouhá léta jako jeden ze základních zdrojů vědeckých informací v humanitních vědách.¹ S nástupem digitálních technologií a zejména s rychlým rozvojem příslušných softwarových nástrojů však získávají nové možnosti využití: přestávají být pouhým zdrojem referencí o existující literatuře či primárních textech, ale nově začínají být využívány jako komplexní datasety, které mohou poskytnout podkladová data pro kvantitativní či statistický výzkum nejen v oblasti humanitních oborů, jako je v našem konkrétním případě literární věda.

Literární věda či jí příbuzné obory v tomto ohledu zauímají specifickou pozici i mezi humanitními disciplínami jako takovými: poněvadž je základním předmětem jejího výzkumu text nebo jeho fyzická reprezentace (kniha, článek či v poslední době samozřejmě též jejich digitální verze), mohou bibliografické databáze sloužit jako vhodný nástroj pro mapování, deskripci a analýzu zvoleného literárního či libovolně jinak definovaného pole. Na základě bibliografických dat totiž lze sledovat nejen „pouze“ cirkulaci či recepci určitého textu, ale lze mapovat i složitější vztahy a vazby mezi jednotlivými účastníky literárního dění, ať již jde o jednotlivé osoby, umělecké skupiny, korporace, periodika, nakladatele a nakladatelství atd., nebo jednotlivá díla, témata či výzkumné směry. V mnoha

¹ Studie vznikla jako výstup projektu Česká literární bibliografie (LM2018136), podpořené ho Ministerstvem školství, mládeže a tělovýchovy České republiky v rámci jeho aktivit na podporu výzkumných infrastruktur.

ohledech se kvantitativní výzkum bibliografických dat může inspirovat existujícími bibliometrickými postupy a nástroji, zároveň však otevírá možnost analýzy značně širšího spektra badatelských otázek a problémů, které nemusejí primárně souviset s mapováním vědeckého provozu a citační metrikou.

Nástup kvantitativního výzkumu v literární vědě je spojován především se jménem Franca Morettiho, jehož některé publikace byly přeloženy též do češtiny,² či jeho spolupracovníků ze Stanford University, resp. Stanford Literary Lab (zejména Mathew Lee Jockers).³ Byť Morettiho práce byly v poslední době opakovaně kritizovány především kvůli nedostupnosti analyzovaných datových souborů a s tím související obtížné verifikovatelnosti jeho závěrů (zejména K. Bode, zde odkazy na další literaturu),⁴ nelze popírat, že kvantitativnímu a statistickému výzkumu v oboru literární vědy či širěji humanitních věd obecně významně napomohly otevřít cestu. Morettiho impulsů, dále stimulovaných rychlým rozvojem výpočetní techniky a softwarových nástrojů, se chopila celá řada následovníků, kteří výsledky jeho bádání dále rozvíjejí či inovativně využívají pro řešení svébytných badatelských problémů zasahujících řadu nejruznějších oblastí literární vědy. Tento trend je zřetelný, i pokud jde o vytěžování bibliografických dat, která jsou předmětem této stati a kterým se v posledních několika letech začíná dostávat zasloužené pozornosti. Jmenujme v této souvislosti alespoň Macieje Maryla a jeho kvantitativní výzkum polské literatury po roce 1989⁵ či Roberta Pétera a jeho bádání o anglické literatuře 18. století a zejména návazný projekt, v jehož rámci vznikl software AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts), na jehož platformě jsou vyvíjeny nástroje právě pro kvantitativní analýzy textových a bibliografických dat.⁶ Zřejmě nejsystematičtější je v současnosti kvantitativní výzkum bibliografických dat provozován při Helsinku Computational History Group (Mikko Tolonen, Leo Lahti

2 F. Moretti: *Grafy, mapy, stromy*; či F. Moretti: *Distant Reading*.

3 M. L. Jockers: *Macroanalysis*. Sledování aktivity Stanford Literary Lab je možné zejména prostřednictvím série online publikovaných tzv. „pamfletů“, viz *Pamphlets*.

4 K. Bode: The Equivalence of „Close“ and „Distant“ Reading, s. 77–106.

5 M. Maryl: Literary Transition in Poland Viewed Through Bibliographical Data (1989–2000). Textová podoba studie je momentálně v tisku – srov. M. Maryl: *Operationalising the Change*, (v tisku).

6 R. Péter – Z. Szántó – J. Seres – V. Bilicki – G. Berend: AVOBMAT, s. 43–55.