

Pochopení učení v online prostředí za pomoci vybraných nástrojů webové analytiky

Webová analytika představuje jednu ze zajímavých oblastí, ze kterých klasické pojetí learning analytics sice vychází, ale současně její metody příliš aktivně nevyužívá. Příspěvek se zaměřuje na využití konkrétních metod, technik a nástrojů webové analytiky, které mohou pomoci objasnit to, jakým způsobem se studenti v online prostředí učí.

Hlavní pozornost bude věnována možnosti využití Google Analytics pro zodpovězení pěti výzkumných otázek: 1) Kdo jsou studenti kurzu a jaké technologie využívají? 2) Jakým způsobem interagují s učebním materiálem? 3) Má design kurzu vliv na způsob interakce s online učebním materiálem? 4) Jaké jsou vzorce chování studentů? 5) Jaké jsou silné a slabé stránky takové analýzy?

Druhá část se bude věnovat aplikaci Smartlook, která umožňuje vytvářet heatmapy stránek a sledovat místa, na kterých uživatelé trávili více času, analyzovat pohyby myši atp. Takto získaná data se pokusíme dát do kontextu s informacemi o úspěšnosti studentů v případě testových otázek. Pokusíme se tak vysledovat souvislost mezi konzumací vzdělávacího obsahu a výsledky testů.

V třetí části se pokusíme podívat na některé další možné analytické nástroje, ať již jde o implementaci vlastních analytických metod do LMS prostředí, klasické learning analytics nebo na analýzu sociálních sítí studentů, což představuje jeden z důležitých zdrojů dat nutných pro lepší pochopení jednotlivých edukačních situací.

Úvod

Využití webové analytiky patří dnes mezi základní pilíře šířeji pojatého konceptu learning analytics, kterou vymezíme níže. Google Analytics, ale také nástroje jako je Hotjar či Smartlook, New Relic, Live Internet, Yandex. Metrica či WordPress Jetpack, představují základní nástroje, které využívá téměř každý webových vývojář. Zatímco u klasického webu je možné relativně snadno nastavit metriky, které budeme sledovat, respektive cíle, kterých chceme dosáhnout, v případě learning analytics jde o proces podstatně komplikovanější. Zatímco například u e-shopu bude metrikou cesta zákazníka ke zboží a platbě, bude možné optimalizovat web pro určitou skupiny uživatelů a budeme chtít dosáhnout vyššího obrátu, tak v případě procesu učení je takový postup podstatně náročnější. Jak určit, které chování studenta je efektivní a které ne? Jak zjistit, co se student naučil? Ostatně i odpověď na otázku, co je učení, respektive jak má probíhat a co je jeho smyslem, je otázkou volby edukačního paradigmatu. Jinak bude odpovídat konektivist, jiným způsobem konstruktivist, pragmatik nebo fenomenologický pedagog.

Volba paradigmatického uchopení edukačního procesu, což je obecně často podceňované téma, má tedy vliv také na to, jakým způsobem a co v procesu učení chceme zkoumat. Optikou webové analytiky je zřejmé, že nejjednodušší by bylo zaměřit se na behavioralistické uchopení studijního procesu. Tím, že nástroje webové analytiky primárně umožňují sledovat chování studenta, tak pokud bychom o něm mohli uvažovat jako o „černé skříňce“ a proces učení reflektovat jen na základě chování (tedy například pohybu na webové stránce, měřením počtu kliknutí, zapojením se do diskusí atp.), bylo by možné říci, že webová analytika představuje zřejmě jeden z nejdokonalejších výzkumných nástrojů pro online vzdělávání.

Také z hlediska didaktického by se behavioralistická teorie jevila jako velice praktická. Tím, že budeme v článku uvažovat nástroje, které nabízejí jistý anonymizovaný pohled na studenty, lze velice pěkně pracovat s analýzou jednotlivých materiálů – opět projevy chování v interakci s ním by byly v centru

zájmu takto paradigmaticky uchopeného výzkumu. V takovém myšlenkovém rámci je celé vzdělávání realizováno v jednom (nebo více, ale učitelem jasně ohraničených a „ovládaných“) prostředí, takže nasazení webové analytiky vytváří vysoce validní model studijního chování.

Současně se ale domníváme, že behavioralistické edukační paradigma s sebou přináší netriviální množství limitů a problémů, které je třeba uvažovat a reflektovat. Je zde přítomná redukce studenta na bytost bez sociálního či emočního kontextu, absence podpory kreativního a kritického myšlení, ale také snaha o vysokou míru standardizace a unifikace vzdělávání, stejně jako snaha o budování bariéry mezi školou (nebo obecně procesem učení) a vnějším světem. Za těchto důvodů považujeme behavioralistické schéma za neudržitelné, jakkoli snad ještě se někdy objevující. Jaké pedagogické paradigma tedy zvolit?

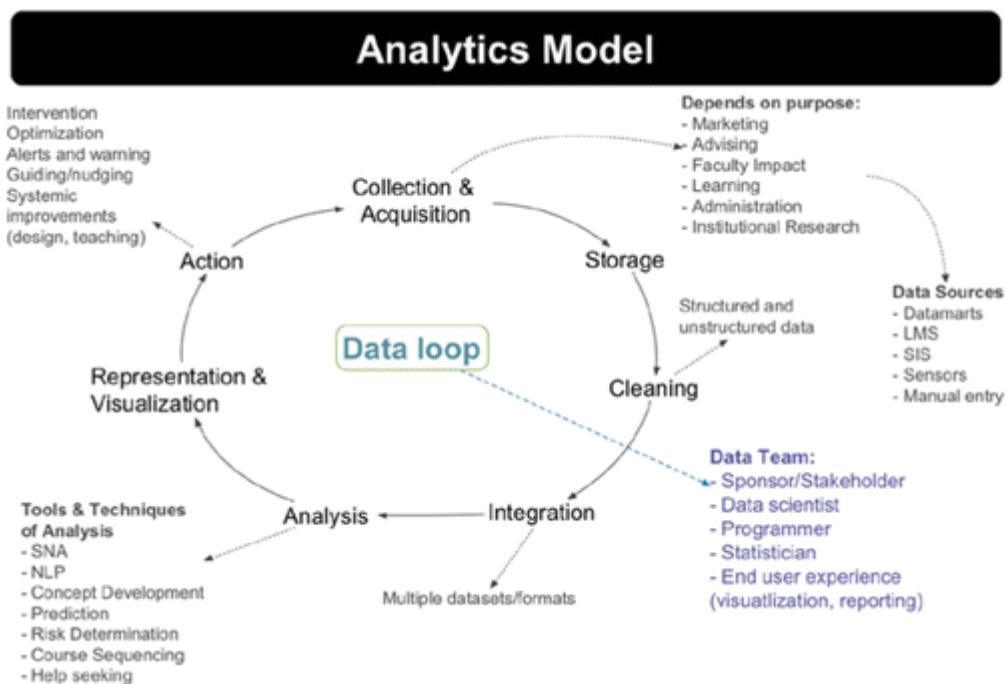
Z čistě praktických důvodů se jako dobrá volba jeví pragmatická pedagogika. Již v článku v minulém ročníku konference jsme ukazovali aspekty, které nás vedou k tomu, že právě programatická pedagogika je efektivním nástrojem v případě online vzdělávání – od pochopitelného jazyka, přes těsný vztah k okolnímu prostředí (funkcemi, jazykem, designem, vzhledem, informační architekturou) až po akcent na užitečnost a praktičnost vzdělávání. Také zde se podržíme pragmatického výzkumného rámce, takže webová analytika nebude moci být jediným zdrojem informací o studentech, ale současně bude mít nějakou vypovídající hodnotu, která může přispět k celkovému pochopení studijního chování daného jedince.

Vymezení pojmu learning analytics není snadné, neboť se lze setkat s různými pojetími a koncepty. Obecně lze říci, že kořeny analýzy učení lze vidět především ve webové analytice (v tomto ohledu se v našem článku vracíme k jejím fundamentálním kořenům); business intelligence, respektive na ni navazující akademickou analytiku; ale také pochopitelně v pozitivistickém a pragmatickém filosofickém paradigmatu; respektive v konceptu (ať již jde o medicínu nebo pedagogiku) vědy založené na důkazech. Dále bychom mohli jmenovat vliv dolování dat, modelování uživatelů a simulací a mnoho dalších, ať již informatických nebo matematických konceptů. Systematický historický vývoj disciplíny zachycuje ve svém článku *Learning Analytics: The Emergence of a Discipline* Siemens (2013).

Lze se setkat s jistým dichotomickým pojetím learning analytics. První možností je, že představuje analýzu vzdělávacího prostředí a procesů v něm. Tento koncept, tedy analytika učení se zabývá měřením, získáváním, analyzováním a prezentací dat o studentech a jejich kontextu za účelem pochopení a optimalizace procesu učení a celého vzdělávacího prostředí, ve kterém edukace probíhá, je spojený především se Siemensem. (Long & Siemens, 2011) Druhou možností je vymezení jejího účelu, jak ji chápe například Ferguson (2012), která chápe learning analytics zaměřené pouze na online prostředí.

Tak jako jsme se z praktických důvodů přiklonili k pragmatické pedagogice, také zde učiníme jisté zjednodušení. Pro pragmatického pedagoga by se jako lepší jevila pozice Siemens. Domníváme se, že pro celkové hlubší porozumění chování studentů je takový přístup nezbytný. V jádru jeho výzkumu je ale obsažena právě analytika práce s online prostředím, tak jak ji chápe Ferguson. Námi zvolený přístup, který budeme reflektovat na konkrétních výzkumných datech tak bude představovat pouze část širšího výzkumu, založeného na smíšeném designu. Na analytická data, která mají spíše kvantitativní charakter, by měly navazovat metody kvalitativní. Pro tento postup máme ještě jeden důvod – na základě kvalitativních dat z analytických nástrojů je možné získat jistý obraz o chování studentů. Pro formování modelu nebo dokonce teorie učení v online prostředí je ale užití kvalitativních nástrojů naprosto nezbytné.

Siemens (2013) pracuje s konceptem analytiky učení v cyklickém pojetí, takže jej vztahuje k modelům, které jsou v obecnější metodologii vnímané jako akční výzkum. Tento proces je zachycen na následujícím schématu, které má s akčním výzkumem spojené také to, že je spojené s intervenčním modelem. Ty mohou směřovat jak k rovině péče o problematické či nadprůměrné studenty (tedy jistě formě krizové intervence), tak také k designu kurzu, optimalizaci prostředí a k dalším činnostem. Zatímco otázka krizové intervence je z hlediska využití Google Analytics problematická, téma designu a optimalizace webu je naopak silně aplikovatelné.



Obrázek1: Siemensův model Learning analytics (Siemens, 2013) je založený na existenci jednotlivých fází získávání a zpracovávání dat, které vedou k provedení jisté akce. Tato činnost je cyklická. Každá změna (ať již na úrovni pedagogické, didaktické, obsahové nebo designové), vede k potřebě nového měření.

Pokud jde o vymezení námi zkoumané oblasti, budeme se zaměřovat na webové kurzy – jde o kurzy, které jsou založené na webu, kterým mohou návštěvníci procházet a učit se z něj. Takové řešení předpokládá, že nedochází k žádné autentizaci studenta. Může tedy jít o studenta univerzitního kurzu (z různých důvodů můžeme předpokládat, že v našem případě tvoří velkou většinu studentů¹), stejně jako náhodní návštěvníci. Způsob pohybu po webu na jednu stranu bude reflektovat obecné zvyklosti studenta, ale jak se ukazuje, velice podstatný bude také rozměr motivační – „je třeba vše přečíst, abych mohl uspět u zkoušky“. Například tento faktor zásadním způsobem ovlivňuje chování studentů na webu.

Námi studovaný kurz je mashup kurz (Auinger, 2009), který kombinuje tři prostředí:

Pro zápis studentů, odevzdávání úkolů a vzájemnou zpětnou vazbu budeme pracovat s univerzitním informačním systémem. Jde také o prostředí pro zápis známek.

¹ Námi studovaný kurz – právě z důvodu výzkumné analýzy je nastavený tak, aby ho neindexovali roboti. Dostávají se tak na něj studenti kurzu, kteří mají odkaz, případě osoby, kterým studenti odkaz poskytnou. Takových osob je ale zřejmě poměrně málo. Lze tedy uvažovat v podstatě o datech, která odrážejí populaci studentů, kteří mají motivaci dokončit kurz a splnit všechny jeho podmínky.

Primární komunikační prostředí je Facebook, respektive skupina na něm. Interakce mezi jednotlivými studenty zde lze měřit běžnými nástroji (Serrat 2017, Scott 2017, Thirumalai et al. 2017). Pokud bychom chtěli systematictěji měřit právě sociabilitu v online prostředí, bylo by možné využít tento nástroj.

Samotný vzdělávací obsah je umístěn na webu. Pro webovou analytiku je možné užít Google Analytics (tak činíme my, ale obecně je možné sáhnout také po dalších nástrojích) a případně po aplikacích jako je Hotjar (Bergs 2017) nebo Smartlook (ten je zvolen v námi sledovaném příkladu, ale funkce jsou s Hotjar relativně podobné).

Možnosti výzkumu a jeho metodologické ukotvení

Na tomto místě bychom rádi nabídli jisté metodologické ukotvení celého výzkumu, popsali výzkumné nástroje a způsob směru dat. Budeme se přitom částečně vymezovat vůči klasickému modelu learning analytics, protože je v literatuře i v běžném pedagogickém diskursu známější a vnímáme jako důležité upozornit na některé případné rozdíly mezi tím, když data zpracované pomocí nástrojů v LMS, tedy můžeme pracovat s každým jednotlivým studentem a mezi obecnou webovou analytikou, kterou využíváme v našem výzkumu.

Pro náš výzkum užíváme Google Analytics a Smartlook, který získává data prostřednictvím Google Tag Manager. Zatímco Google Analytics poskytuje informace, které umožňují provést deskripci celé populace a nabídnout pohled na využívané technologie, tak Smartlook je zaměřen spíše na práci s chováním jednotlivých uživatelů na stránce nebo na analýzu interakce uživatelů s konkrétní webovou stránkou.

Obecně je možné oba nástroje použít ve dvou samostatných kontextech, které pracují se stejnou množinou dat, ale nabízejí jiný přístup:

- a) Lze pracovat s anonymizovanou neadresnou deskripcí – tedy věnovat se vztahů, událostem a datům, která vypovídají o celkovém chování populace. Jde o základní výzkumný přístup, který budeme více sledovat také v tomto textu.
- b) Jak Google Analytics, tak také Smartlook nabízejí možnost sledovat chování jednotlivého uživatele. Oba nástroje provádějí jeho anonymizaci, která není textově synchronní, ale obsahuje časové značky. Pomocí nich je pak možné oba výzkumné nástroje propojit a použít pro kvalitativní výzkumné metody.

Pokud jde o výzkumný vzorek, je možné pracovat v oblasti vzdělávání se dvěma možnými přístupy. Tím prvním je práce s indexací otevřenými webovými kurzy, ve kterých je výzkumný vzorek nejasně ohraničen. V takovém případě je možné využít pro jeho popis demografická data z Google Analytics a pomocí nich identifikovat místo připojení nebo přibližný věk. Hlubší deskripce takto koncipovaného vzorku není snadná, ale na mnoho údajů lze případně usuzovat také z analýzy zdrojů návštěvnosti. Touto metodou lze například odhadnout, kolik je v populaci skutečných „kurzových studentů“ a kolik „kolemjdoucích návštěvníků“.

Druhou možností, kterou většinou využíváme je kurz v aktuálním běhu nenechat indexovat roboty. Výsledkem je prakticky dokonalý vzorek uživatelů, kteří kurz buď sami přímo studují, nebo někomu museli poskytnout odkaz. Ukazuje se, že tato návštěvnost mimo zamýšlený vzorek „studentů“ je zcela zanedbatelná. Díky tomuto opatření získáme informace o vzorku, které můžeme čerpat ze školního informačního systému (takže známe například obory studia, ročník, pohlaví,...) a další deskriptory, které mohou pomoci s interpretací výsledků. Platí, že nemůžeme snadno propojit například roční studia s konkrétním chováním uživatele na stránce, ale můžeme z dat usuzovat na jisté chování celé populace.

Data z našeho výzkumu budeme prezentovat na studentech kurzu Kreativní práce s informacemi, který měl (k 7. 5. 2018) 167 studentů – 56 v prvním ročníku bakalářského studia, 28 navazujícího magisterského studia, 26 v druhém ročníku bakalářského studia a 25 v druhém ročníku navazujícího magisterského studia. Zbytek studentů byl ve vyšších ročnících. Kurz studovalo 56 mužů a 113 žen. 84 z nich studovalo na filosofické fakultě, 26 na fakultě informatiky, 23 na ekonomicko-správní fakultě, zbytek, tedy 36 studentů na dalších fakultách Masarykovy univerzity. Tyto statistické údaje pocházejí z informačního systému MU. Podle dat Google Analytics (z doby 23. 4. - 6. 5) bylo evidováno 108 uživatelů ve věku 18-24 bylo 71 osob a 23-34 33 osob. Jiná věková skupiny zaznamenána nebyla. Žen bylo 83 a mužů 27.

Etická stránka výzkumu je ošetřena informací o tom, že na stránce se užívají cookies, Google Analytics a Smartlook pro webovou analýzu. Data z takto získaných zdrojů je možné vnímat jako anonymizovaná, jakkoli se samozřejmě může stát, že například uživatel ze specifického místa (například konkrétního státu mimo ČR) bude tímto způsobem „prozrazen“. Dle našich měření jde ale o poměrně vzácný jev a reálně není možné mu dostatečně efektivně zabránit. V kontextu běžně chápané analytiky učení, kde může (například v Moodle(Dimopoulos 2013)) učitel detailně sledovat učební chování studenta, ale také například to, v kolik hodin se do systému připojuje, je možné říci, že námi prováděný výzkum je v rámci learning analytics jeden z nejvíce citlivých k soukromí studenta. To souvisí také s tím, že náš výzkum neumožňuje cílenou paternalisticky koncipovanou intervenci směrem ke konkrétní osobě, která by vykazovala zvláštní struktury svého studijního chování (například pozdě odevzdávala úkoly, málo času trávila předepsanými aktivitami apt.).

Oba výzkumné nástroje, které používáme pro účely našeho textu, jsou primárně nástroji na webovou analytiku, nikoli na analýzu učení, což s sebou přináší specifické problémy, ale také možná nové možnosti. Domníváme se, že práce s datovou analytikou, ale také s metrikami a nástroji, které pro měření může používat, nabízí netradiční a do jisté míry tedy prohlubující rozměr studia chování jednotlivých účastníků.

V rámci našeho výzkumu budeme tedy kombinovat data získaná dvěma měřícími nástroji – Smartlook a Google Analytics(Ledford 2011, Omidvar 2011; Pakkala 2012; Romanowski 2016). Smartlook je českou obdobou hotjar a umožňuje především dvě činnosti, které jsou pro nás významné. Tou první je práce se záznamy – aplikace nahrává pohyb uživatele po stránce a na základě jeho IP adresy ho umí anonymizovat, ale současně spojit s jedním unikátním jménem. V systému je tedy možné sledovat, jak vypadá (anebo se mění) chování návštěvníka v průběhu času, zda transformuje své chování v závislosti na délce studia kurzu nebo podle zvoleného tématu a jeho délky, jakým způsobem se mění jeho chování při interakcích s různými médii apt. Druhou funkcí, kterou hotjar nabízí je tvorba heatmap na webových stránkách. Ty umožňují například sledovat místa, na která uživatelé více klikali nebo kolem kterých kroužili myši. Lze také sledovat, do jaké části stránky uživatelé srolovali, nebo rozdíl mezi telefonem, tabletem a desktopem. Smartlook je tedy možné užít především pro kvalitativní analýzu. (Černý 2018)

Na tomto místě musíme upozornit na omezení, které je v našem ukázkovém vzorku spojené s ekonomickými možnostmi výzkumu. Heatmapy je možné konstruovat pouze z jednoho tisíce návštěv (pro více návštěv by bylo třeba mít dražší verzi aplikace). V praxi ale k velkému zkrácení nedochází, protože průměrná návštěvnost je mezi 800 – 1100 návštěvami. Tím, že jsou brány první návštěvy, tak je možné, že námi analyzovaná data nezachycují ty, kteří se stránkou pracují na poslední chvíli, ale vzhledem k charakteru analýzy nejde o závažný problém.

Google Analytics patří mezi nejnámější nástroje pro webovou analytiku vůbec. Pro většinu uživatelů představuje bezplatný nástroj, který umožňuje provádět poměrně rozsáhlé výzkumy přímo na

stránce. Od základních informací, jako je počet návštěv, poměr nových a vracejících se návštěvníků, průměrného času stráveného návštěvníkem nebo na stránce atp., až po poměrně zajímavé informace například o místě připojení, věku uživatelů či užívaným technologiím. Zajímavou informací mohou být ale například také data o využívané síti. Tím, že námi studovaný kurz je univerzitní, lze specificky pracovat se skupinou uživatelů, kteří studují z univerzitního prostředí a případně sledovat jejich odlišnosti od těch, kteří studují jinde. Tato data je možné zkoumat jak na příkladu celé populace, tak také provádět specifické filtrování, až na úroveň jednotlivců, kteří jsou opět anonymizováni.

Google Analytics nemá v našem případě žádný limit z hlediska počtu uživatelů. Data z něj je možné exportovat do CSV, Google Tabulek nebo XLSX, ale také systém podporuje jednoduché kontingenční tabulky. Díky nim je možné snadno nacházet základní vztahy mezi různými proměnnými, aniž bychom museli provádět složitější statistické operace.

Oba měřicí kódy jsou do webové stránky tvořené v systému Umbraco (Wahlberg 2011) vloženy skrze redakční rozhraní. Relevantní otázkou je úplnost a spolehlivost dat. Podle výsledků měření, ale také literatury se zdá, že by data měla být poměrně kompletní. Existují postupy, jak oba měřicí kódy blokovat, ale není možné rozumně předpokládat, že by šlo o jev přesahující jednoho či dva uživatele. Už jen proto, že takový postup je technicky poměrně náročný. Data mohou vykazovat jistou nekonzistenci ve vztahu k osobě – pokud se uživatel připojuje z více zařízení, systém ho identifikuje jako více osob, což znesnadňuje lepší interpretaci některých vzorců chování či studijních návyků. Problematictější může být identifikace místa nebo sítě v případě, že uživatel pracuje s VPN. V takovém případě jsou oba údaje zkreslené. Opět ale neočekáváme, že by šlo o nějak masový jev, byť zřejmě podstatně pravděpodobnější, než blokování Google Analytics na straně uživatele.

Pokud nebylo uvedeno jinak, budeme pracovat s daty z kurzu Kreativní práce s informacemi, který se učí v jarním semestru 2018 (tedy z daty od 10. 2. – 8. 5. 2018). Nemáme v úmyslu zde představovat výzkumnou případovou studii, ale spíše na základě konkrétních dat ukázat, jak je možné na vznesené výzkumné otázky odpovědět, s jakými daty a jakým způsobem je možné je interpretovat.

Výzkumné otázky

Jak již bylo předesláno v abstraktu článku, svoji hlavní pozornost zaměříme na výzkumnou otázku: Jak se studenti v online kurzu chovají. Na základě článku od Luo (2015) jsme identifikovali pěti vedlejších výzkumných otázek, na které bychom rádi odpověděli:

- 1) Kdo jsou studenti kurzu a jaké technologie využívají?
- 2) Jakým způsobem interagují studenti s učebním materiálem?
- 3) Má design kurzu vliv na způsob interakce s online učebním materiálem?
- 4) Jaké jsou vzorce chování studentů?
- 5) Jaké jsou silné a slabé stránky takové analýzy?

V námi provedené analýze bude spojovat možnosti, které nám poskytují oba výzkumné nástroje pro datovou analytiku, abychom ukázali co možná nejkompaktnější pohled na analýzu učení a chování studentů.

Kdo jsou studenti a co o nich víme?

Pokud bychom se zaměřili pouze na data z Google Analytics, tak lze pracovat především se dvěma zdroji informací – tím prvním je pohlaví a věk uživatelů, tedy jejich základní demografický popis. Obecně jde o data z nižší mírou spolehlivosti, než jiná. Druhým zdrojem informací mohou být údaje o místě, ze kterého se připojují. Lze tak například snadno analyzovat, kolik osob se připojuje z území mimo město, kde výuka probíhá nebo například z vesnice. Tyto informace mohou být důležité například pro konstrukci úkolů, kde se můžete ukázat jako problematické sahat k různým formám

blended learning nebo chtějí po uživateli, aby například navštívili knihovnu nebo jiné ve fyzickém prostoru ukotvené místo. Lze také časově škálovat místa připojení a sledovat, kdy a zda se více uživatelů připojuje z prostoru mimo domovské město univerzity.

Jiným zdrojem informací může být informace o počítačové síti. Zde lze relativně snadno filtrovat uživatele, kteří se připojují prostřednictvím školní sítě a ty, kteří studují odjinud. Lze si tak například klást otázku, zda školní prostředí zkracuje nebo naopak prodlužuje dobu studia. To, na co ale webová analytika neumí odpovědět je otázka proč.

Z geolokačních dat lze také snadno identifikovat studenty, kteří studují mimo Českou či Slovenskou republiku a sledovat jejich vzorce chování.

V obecném případě, který ale v námi sledovaném kurzu z důvodu absence indexace obsahu roboty nebo prolinkovaností na weby nenastává, je možné také zjistit, jaká klíčová slova uživatele na web přivádějí (tedy předpokládat jejich zájem). Pokud není indexace zapnutá, zbývá možnost využít výsledky z vyhledávání, které se ukládají jako samostatné webové stránky, jenž jsou pak vidět ve výsledcích návštěvnosti. Otázkou ale je, zda taková data skutečně odpovídají zájmu uživatelů a nikoli špatně nastavenému designu webu, kdy některé pojmy nebo témata nemusí být logicky strukturovaná (vinou informační architektury) a uživatelé je proto hledají. Je zde ještě možnost, že hledají informace, které se objevují v testu. V tomto ohledu je analýza vyhledávání sice důležitá a užitečná, ale její interpretace je značně složitá.

Podobně v našem případě nemůžeme využít ani informace o tom, z jakých stránek do našeho kurzu uživatelé přicházejí, ani jaké využívají vyhledávače. Všechna tato data by bylo ale možné v obecném případě využít.

Zcela zásadní informace jsou ale ty, které se věnují zájmům uživatelů. Google Analytics umí tyto zájmy identifikovat kvůli práci s reklamou, ale pro vzdělávání mají tato data zásadní význam – lze podle nich pracovat s příklady či vymýšlet úkoly, které budou pro danou cílovou skupinu nejzajímavější. V námi studovaném vzorku máme data bohužel až od 23. března. V tabulce uvádíme zájmy, které vykazují nejdelší návštěvnost, tedy předpokládaný nejvyšší studijní zájem:

	Uživatelé	Návštěvy	Prům. doba trvání návštěvy	Počet stránek na 1 návštěvu
Travel/Travel Buffs/Family Vacationers	12(0,25 %)	54(0,17 %)	18:01	8,74
Media & Entertainment/TV Lovers/TV Comedy Fans	15(0,32 %)	54(0,17 %)	17:47	9,06
Lifestyles & Hobbies/Thrill Seekers	16(0,34 %)	75(0,23 %)	17:31	9,00
Travel/Travel Buffs/Beachbound Travelers	16(0,34 %)	85(0,26 %)	16:40	7,35
Travel/Travel Buffs/Snowbound Travelers	17(0,36 %)	122(0,38 %)	15:22	7,80
Media & Entertainment/Music Lovers/Country Music Fans	27(0,57 %)	195(0,60 %)	14:50	7,30
Lifestyles & Hobbies/Pet Lovers/Dog Lovers	20(0,42 %)	103(0,32 %)	13:54	7,58
Media & Entertainment/Music Lovers/Jazz Enthusiasts	15(0,32 %)	71(0,22 %)	12:46	7,03
Lifestyles & Hobbies/Business Professionals	45(0,95 %)	298(0,92 %)	12:42	6,90
Travel/Business Travelers	40(0,85 %)	274(0,84 %)	12:37	6,67

V případě, že bychom chtěli sledovat nejčastější zájmy uživatelů, pak by pořadí bylo následující:

	Uživatelé	Návštěvy	Počet stránek na 1 návštěvu	Prům. doba trvání návštěvy
Media & Entertainment/Music Lovers	194	1450	6,26	10:29
Media & Entertainment/Book Lovers	189	1283	6,13	10:17
Lifestyles & Hobbies/Art & Theater Aficionados	188	1433	6,18	10:39
Media & Entertainment/Movie Lovers	183	1281	6,38	10:56
News & Politics/News Junkies/Political News Junkies	169	1171	6,05	10:17
Sports & Fitness/Health & Fitness Buffs	163	1208	6,38	10:47
Lifestyles & Hobbies/Green Living Enthusiasts	159	1212	6,09	10:36
Lifestyles & Hobbies/Pet Lovers	149	978	6,19	9:50
Food & Dining/Foodies	129	987	6,28	11:05
Lifestyles & Hobbies/Shutterbugs	121	833	6,11	10:19

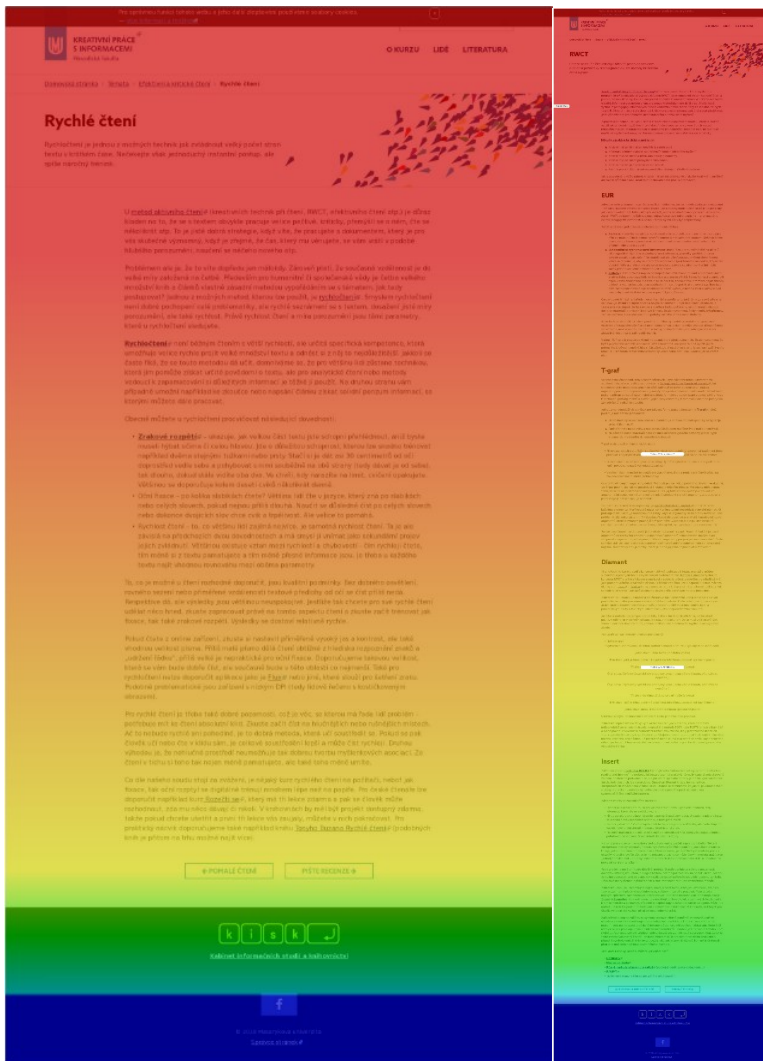
Průměrné hodnoty jsou 6,32 stránek na návštěvu a 10:49 minut strávených jednou návštěvou. Google Analytics umožňuje ještě další strukturalizaci kategorií zájmu uživatelů – obecně by tato data dávají dobrý, byť zřejmě ne zcela přesný pohled na to, jací uživatelé se v kurzu pohybují a představují jeden nástroj pro efektivnější design nebo pro práci s emočním designem kurzu.

Samostatnou otázkou je, jaké technologie využívají studenti daného kurzu, což je také součástí jejich profilu. Této problematice se ale věnujeme v samostatné výzkumné otázce.

Jakým způsobem interagují studenti s učebním materiálem?

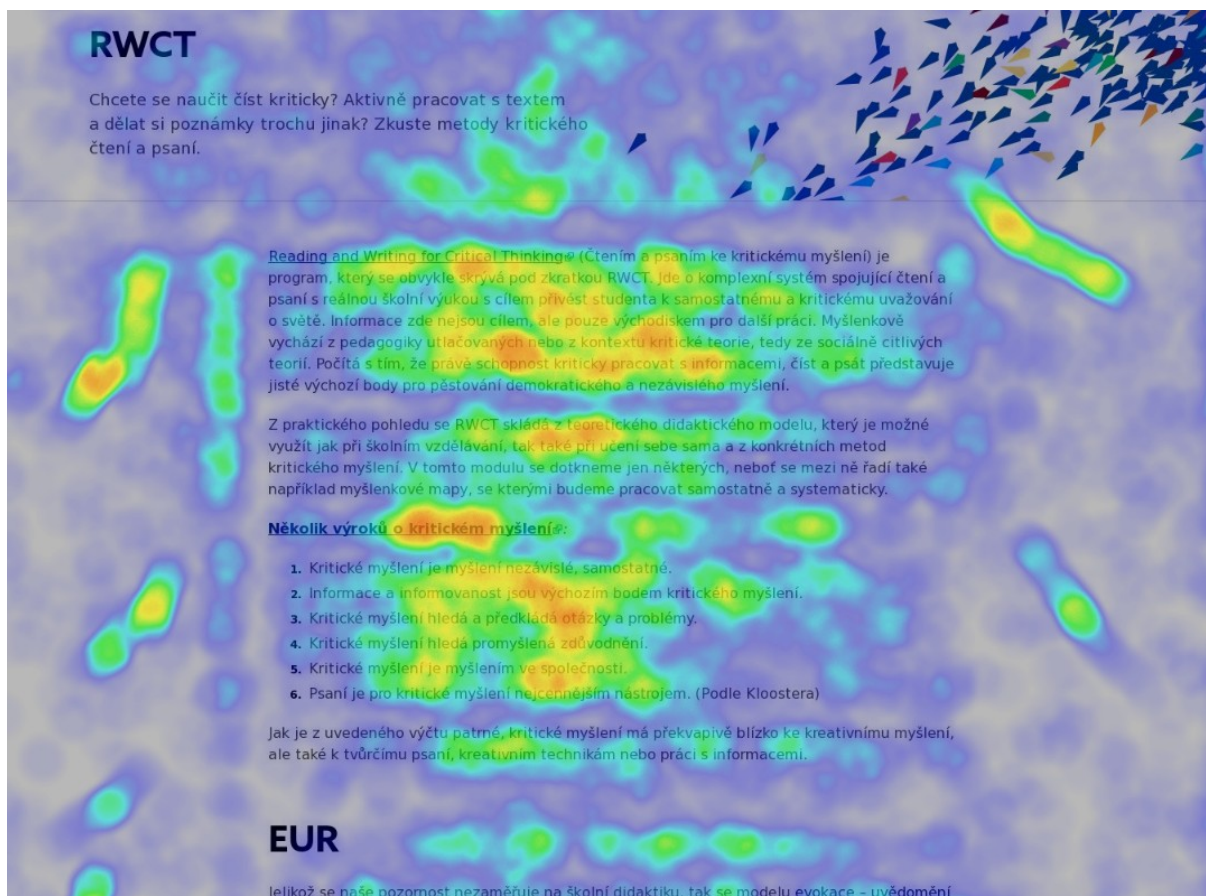
Výzkumná otázka „Jakým způsobem interagují studenti s učebním materiálem?“ těsně souvisí jak s otázkou po technologiích, které používají, tak také s otázkou po vzorcích chování. Její částečné zodpovězení tedy ponecháme do dalších kapitol a zde se zaměříme jen na některé specifické aspekty interakce studentů s obsahem.

Z dat dostupných ze Smartlook se jeví jako zajímavé také informace o tom, kolik uživatelů skroluje na stránce jak daleko. Má například délka stránky vliv na to, kolik uživatelů se podívá až na konec?



Vpravo na obrázku je ilustrativně zvolený příklad poměrně krátké stránky v kurzu v levé části pak stránky mimořádně dlouhé. Je patrné, že pokud chceme, aby studenti texty četli, je třeba tomu přizpůsobit také délku textu. Dlouhé texty mají nižší míru dočtenosti až do konce.

Tak, jak jsme naznačili již výše, základními parametry pro analýzu jsou také heatmapy, které umožňují identifikovat místa na webu, která mají nějaké specifické struktury uživatelské interakce. Typicky umožňují identifikovat místa, na která uživatelé klikají nejčastěji, nebo kde se nejvíce pohybují kurzorem, což odkazuje k místům, která jsou zajímavá či náročná nebo naopak lze najít oblasti na stránce, které mají míru zajímavosti podstatně nižší.



Obrázek ukazuje analýzu pohybu myši na konkrétní stránce. Studenti věnují pozornost konkrétnímu výčtu a pak v prvním odstavci pojímům „kritická teorie“ a „pedagogika utlačovaných“, které jim nejsou jasné a známé. Pochopitelný je také zájem u odkazu, který vede na povinný studijní materiál.

Pokud jde o analýzu proklikovosti, pak lze užít jednak nástroj Google Analytics Tok chování, který umožňuje ukázat, jakým způsobem se uživatelé přepínají mezi jednotlivými stránkami. Tomuto nástroji se budeme podrobněji věnovat dále. Pracovat je možné ale také s doplňkem pro Google Chrome, který umožňuje zobrazovat na stránce informace o klání na interní odkazy. Užitečným nástrojem v této oblasti je také samotný Smartlook, který umožňuje identifikovat výřez na stránce (ať již jde o odkaz nebo skupinu odkazů) a kvantifikovat počet kliknutí. Pomocí Google Analytics lze také sledovat stránky, na které studenti odcházejí, což jsou typicky externí odkazy.

V námi studovaném kurzu se příliš nedají využít data z vyhledávání, takže se zdá, že studenti v kurzu postupují poměrně značně lineárním způsobem. Ostatně i na datech z Kurzu práce s informacemi, který byl stavěný od počátku jako nelineární (respektive s možností nelineárního studia) se velice snadno a jasně ukázalo, že studenti výrazně preferují posloupnost materiálů. Jestliže nějakým způsobem „přeskakují“, pak zpravidla tak, že výrazně preferují informace o testu a úkolu před vlastním edukačním obsahem.

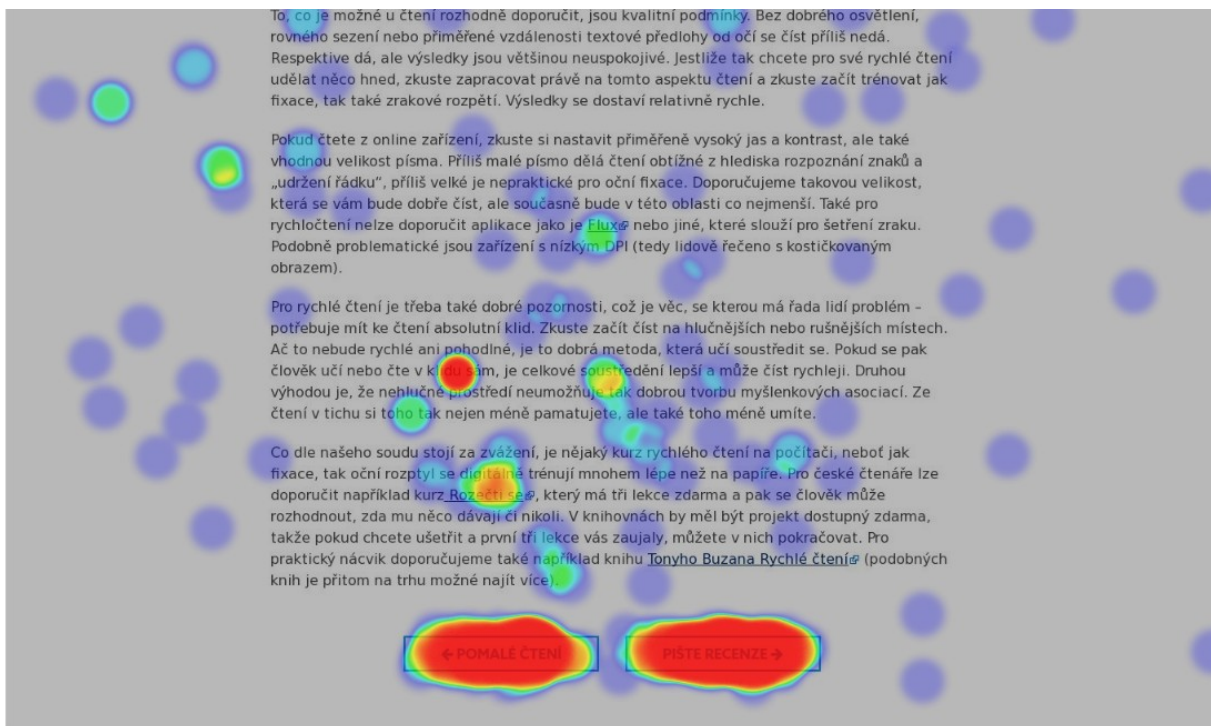
Má design kurzu vliv na způsob interakce s online učebním materiálem?

Na tomto místě se bohužel nemůžeme opřít o relevantní výzkumná data, která by bylo možné získat A/B testováním, což je výzkumná metoda, pro kterou má Google k dispozici také specifické nástroje. Proto si dovolíme jen několik poznámek, které shrnují zkušenosti spíše praktického než výzkumného charakteru.

Sesterský Kurz práce s informacemi probíhal v semestrech podzim 2017 a podzim 2016 ve dvou vizuálně odlišných prostředích, ale s poměrně značně podobným obsahem, podmínkami a strukturou. V tomto kurzu se ukázalo jako významné, jak kurz vypadá. Příliš se nezměnili základní charakteristiky studia – čas na jednu návštěvu byl v roce 2017 8:28, počet stránek 5,07 a v roce 2016 8:06 a 5,13, takže jde v podstatě o totožná čísla. To, co se významně lišilo, byly stránky, které uživatelé navštěvovali, takže například organizační pokyny byly v roce 2006 poměrně často čteny a 2017 již nikoli. Tuto skutečnost lze také přisoudit tomu, že se změnil formát komunikace se studenty.

Zajímavé je pak srovnání kurzu Kreativní práce s informacemi a Kurzu práce s informacemi v roce 2017, kdy byla užita stejná struktura i stejný vizuální styl. Počet stránek na jednu návštěvu v průměru v Kreativní práci s informacemi byl 6,19 a čas 10:22, což jsou výrazně vyšší čísla než v podzimním kurzu. Otázkou je, jaké má tento stav příčiny – lze se domnívat, že pokud má Kreativní práce s informacemi čtrnáctidenní periodu s novými materiály, mohou ji studenti věnovat studiu více času.

Pro přesnější měření a obecnější závěry by bylo ale třeba postupovat exaktněji. Snad jen poslední poznámkou může být fakt, že mezi kapitolami je možné přecházet pomocí tlačítek „vpřed“ a „zpět“ umístěných na konci textu. Jednoznačně se ukazuje, že studenti s tímto linearizačním prvkem aktivně pracují.



Heatmapa ukazující kliknutí na tlačítka navigace. Zajímavé je, že tlačítko „zpět“ (225) má jen o relativně málo kliknutí méně než to, které odkazuje na další stránku (308). Lineární učení tedy z tohoto obrázku (ale také z dalších dat z jiných stránek) nepůsobí jako čistě dopředné, ale spíše jako „nepřeskakující“.

Jaké jsou vzorce v chování studentů?

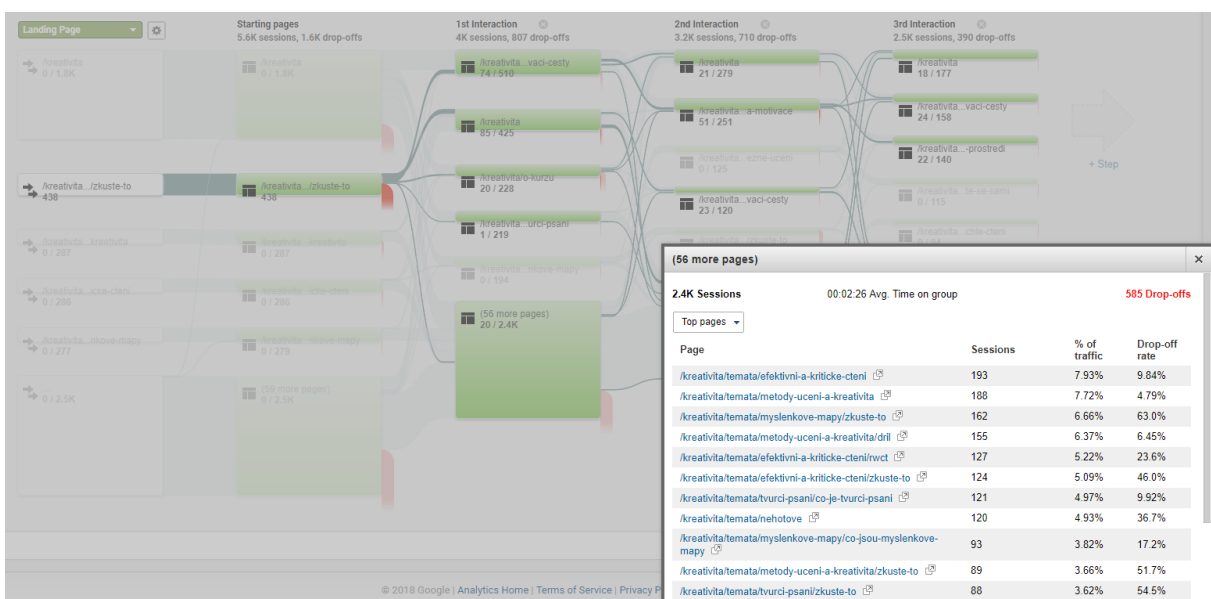
Čtvrtá výzkumná otázka zní: „Jaké jsou vzorce v chování studentů?“ Zde je třeba říci, že jde o otázku nesmírně komplexní a není možné na ni jednoduše odpovědět. Spojení Smartlook a Google Analytics v této oblasti nabízí velké množství oblastí, kterým se lze výzkumně věnovat. Zde je třeba zdůraznit, že námi nabízený pohled není ani zdaleka úplný.

Základními metrikami, které je možné sledovat jsou v Google Analytics navštívené stránky a doba, kterou student na stránce stráví. Díky těmto údajům je možné pracovat například s předpokládanou rychlostí čtení student, s analýzou doby, kterou student celkově v kurzu stráví a s řadou dalších důležitých údajů.

Velice užitečným, byť na analýzu náročnějším nástrojem je mapa toku uživatelů. Umožňuje sledovat chování uživatelů na webu z hlediska jejich průchodu službou – jaké jsou jejich typické cesty ke zdroji, z jakých stránek odcházejí a kam naopak přicházejí, jaký je jejich typický pohyb prostředím? Tato data umožňují sledovat jak kvalitu a strukturu webu a porovnávat je s předpokládaným průchodem. Jde o data, která mohou velice významně pomoci se sledováním návazností v kurzu, ale také významně pomoci se studiem chování uživatelů.

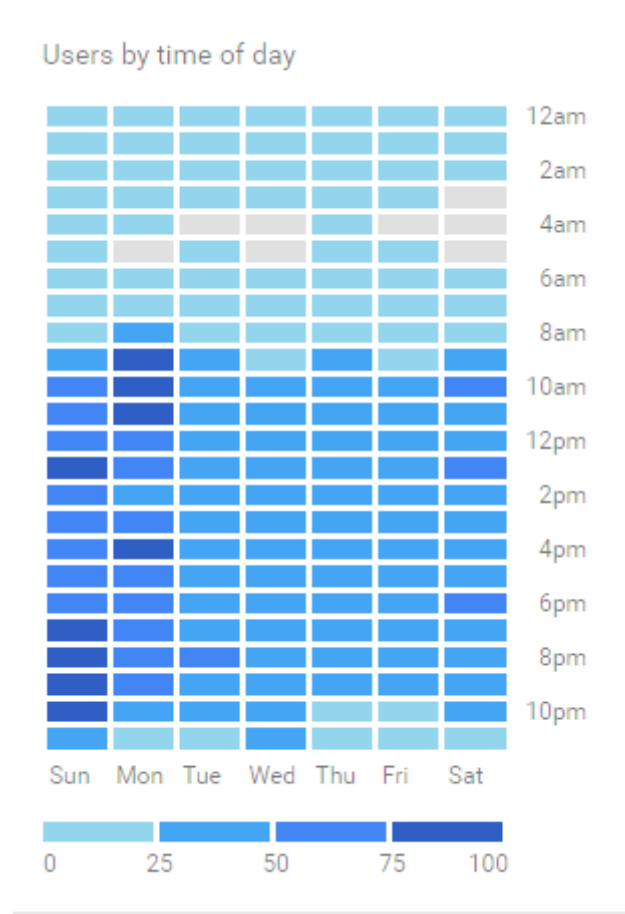


Tok uživatelů zachycuje, jakým způsobem se uživatelé pohybují na webu, odkud a kam přecházejí a kde jsou místa, kde web opouštějí. Na obrázku je zvýrazněné chování uživatele, který se nachází na jedné konkrétní stránce.



Na tomto obrázku jsou ukázány informace o pohybu uživatelů z konkrétní stránky v kurzu – lze si tak vytvořit poměrně přesný a adekvátní model pohybu studentů z konkrétních míst na webu.

Tok uživatelů představuje velice zajímavá data o chování uživatelů, které není možné žádnou jinou běžně dostupnou analýzou získat. (Tonyan 2015; Náповěda Google Analytics, 2018)² Porozumění tomu jak studenti studují, procházejí webem, odcházejí z něj a kolik kde tráví času, představuje zajímavá data, která opět mohou výrazně pomoci s popisem studentského života a chování. K těmto datům lze přičíst také analýzu toho, v jakém čase studenti web navštěvují, případně jaký je rozdíl mezi mobilní a desktopovou návštěvností.



Čas a den v týdnu, kdy studenti navštěvují web. Do nedělní půlnoci je třeba odevzdat domácí úkol, proto je zde návštěvnost nejvyšší. Výše vizualizovaná data jsou získaná za 90 dní. Pondělní zvýšená aktivita je dána tím, že v pondělí odchází informační mail studentům se zadáním úkolu nebo otevření tématu. Jinak je zajímavé, že kurz nabízí velice rovnoměrné rozložení studia v čase od úterý do soboty a od 10 AM do 10 PM. Čas mezi 3 AM a 5 AM je logicky nejméně využíván.

Smartlook umožňuje pracovat se záznamy pohybu uživatelů na webu, takže výzkumník může využít metody analytiky, které jsou všeobecně známé například s práce s videm ve výuce (Santagata 2007; Goldman 2007; Minačíková 2016; Janík 2013). Díky této analýze je možné vytypovat konkrétní uživatele (v případě velkého výzkumu i celý výzkumný vzorek) a jejich chování v hodným způsobem kódovat. Tím je možné – v kontextu předchozích dat – získat poměrně komplexní informace o chování studentů v průběhu studia.

². Časté jsou také texty A/B testování v kontextu toho nástroje (Young, 2014, 2014a)

Jaké jsou silné a slabé stránky takové analýzy?

Pokud jde o silné stránky takové analýzy, lze říci, že dochází ke studiu chování studenta v „přirozeném prostředí“ jeho studia. Nejde tedy o práci ve speciálním LMS, kde mohou být studijní činnosti značně odlišné od běžného chování se na webových stránkách, ale skutečně o sledování toho, jak student pracuje. V tomto ohledu se domníváme, že využití nástrojů webové analytiky nabízí širší validnější výsledky než běžná learning analytics.

Za cenné také považujeme to, že můžeme pečlivě analyzovat interakce studentů se studijními materiály. Oproti běžné práci s learning analytics lze anonymizovaně sledovat práci studenta, to, co na stránce hledá a jak pracuje. Také se lze dostat k poměrně dobrým odhadům zájmů studentů nebo k jejich specifickým modelům chování. Za zajímavé lze považovat také to, že lze dobře sledovat rozdíly mezi učením se na tabletu či v mobilním telefonu a na desktopu.

V klasickém paradigmatickém rámci learning analytics lze považovat za největší nevýhodu to, že webová analytika neumožňuje provádět žádné intervence. Nelze pomoci studentovi, který prokazuje jisté problematické chování, neboť není možné ho adresně identifikovat. S tím souvisí také to, že nelze snadno provázat tyto metody například s rozhovory či jinými metodami, které by předpokládali práci s jednotlivými uživateli. Důvodem je právě nemožnost snadného provázání jednotlivých aktérů vzdělávání s daty. Z hlediska relevance je problematické také to, že pokud student využívá více IP adres, což je typické, není možné o jeho chování vypovídat s dostatečnou konzistencí – data jsou principiálně neúplná a výzkumník neví jakým způsobem. Například tvorba různých konceptuálních modelů je z tohoto důvodu problematická.

Závěr

V článku jsme se pokusili identifikovat základní metodologický postup, který umožní pracovat s metodami webové analytiky jako s nástroji analýzy učení. Pro tyto účely jsme se rozhodli nabídnout stručnou ilustraci tématu na datech pocházejících z právě běžícího kurzu. Jeho obsáhlejší a širší analýza formou případové studie bude předmětem naší publikační činnosti, hned jak bude kurz ukončen a budeme moci pro jeho reflexi užít také data z testů nebo zpětných vazeb od studentů.

Domníváme se, že využití Google Analytics i Smartlook umožňuje hlubší a metodologicky odlišné zamyšlení se nad tím, jak studenti v online prostředí studují a jakým způsobem interagují s učebními materiály. Doufáme, že předložený text může posloužit jako základní metodologické východisko pro případné další výzkumy nebo jako pomoc pro designery kurzů. Právě tvůrci online vzdělávání by s daty o uživateli měli pracovat co možná nejvíce a nejpečlivěji, protože na základě nich lze nabídnout efektivnější a adekvátnější vzdělávací obsah.

Literatura

Auinger, A., Ebner, M., Nedbal, D., & Holzinger, A. (2009). Mixing content and endless collaboration–MashUps: Towards future personal learning environments. *Universal access in human-computer interaction. Applications and services*, 14-23.

Bergs, H. (2017). *Gaining insights into website traffic and online consumer behavior with Google Analytics* (Doctoral dissertation).

Černý, M. (2018). Využití nástrojů webové analytiky pro pochopení učení v online prostředí. *ProInflow: Časopis pro informační vědy*, 10(1).

Dimopoulos, I., Petropoulou, O., Boloudakis, M., & Retalis, S. (2013). Using Learning Analytics in Moodle for assessing students' performance.

- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317.
- Goldman, S., & McDermott, R. (2007). Staying the course with video analysis. *Video research in the learning sciences*, 101-113.
- Janik, T., Minarikova, E., & Najvar, P. (2013). Der Einsatz von Videotechnik in der Lehrerbildung. Eine Übersicht leitender Ansätze. *Videobasierte Kompetenzforschung in den Fachdidaktiken*, 63-78.
- Ledford, J. L., Teixeira, J., & Tyler, M. E. (2011). *Google analytics*. John Wiley and Sons.
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, 46(5), 31–40.
- Luo, H., Rocco, S., & Schaad, C. (2015, October). Using Google Analytics to Understand Online Learning: A Case Study of a Graduate-Level Online Course. In *Educational Innovation through Technology (EITT), 2015 International Conference of*(pp. 264-268). IEEE.
- Minačíková, E., Pířová, M., & Janík, T. (2016). Using VideoWeb in EFL Teacher Education: do the benefits differ between teachers with and without previous teaching experience?. *The Career Trajectories of English Language Teachers*.
- Omidvar, M. A., Mirabi, V. R., & Shokry, N. (2011). Analyzing the impact of visitors on page views with Google analytics. *arXiv preprint arXiv:1102.0735*.
- Pakkala, H., Presser, K., & Christensen, T. (2012). Using Google Analytics to measure visitor statistics: The case of food composition websites. *International Journal of Information Management*, 32(6), 504-512.
- Romanowski, B., & Konak, A. (2016). Using Google Analytics to Improve the Course Website of a Database Course.
- Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of mathematics teacher education*, 10(2), 123-140.
- Serrat, O. (2017). Social network analysis. In *Knowledge solutions* (pp. 39-43). Springer, Singapore.
- Scott, J. (2017). *Social network analysis*. Sage. Thirumalai, C., Sree, K. S., & Gannu, H. (2017, April). Analysis of cost estimation function for facebook web click data. In *Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of* (Vol. 2, pp. 172-175). IEEE.
- Siemens, G. (2013). Learning Analytics [Online]. *American Behavioral Scientist*, 57(10), 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Wahlberg, N., & Sterling, P. (2011). *Umbraco User's Guide*. John Wiley & Sons.
- Young, S. W. (2014). A/B Testing Web Analytics Data [dataset].
- Young, S. W. (2014b). Improving library user experience with A/B testing: Principles and process. *Weave: Journal of Library User Experience*, 1(1).