

Testování použitelnosti v kartografii a geoinformatice

Herman, Lukáš

O čem bude řeč?

- 1) Úvod a motivace,
- 2) Míry a ukazatele v rámci studií použitelnosti
- 3) Nejvýznamnější používané metody testování a hodnocení použitelnosti,
- 4) Realizace experimentů a jejich popis,
- 5) Statistické zpracování,
- 6) Dotazy a diskuse.

Proč? K čemu mi to bude dobré?

Několik úryvků ze zadání Vašich diplomových prací:

... verifikace vytvořených kartografických vizualizací..., ...uživatelské ověření..., ...posoudit vliv na efektivitu..., ...validace výsledků..., ...analýza vztahů..., ...vyhodnocení dosažených výsledků statistickými metodami..., ...podrobeny praktickému testování za účelem jejich optimalizace..., ...analýza závislosti..., ...analýza souvislostí..., ...srovnávat data..., ...vlastní výsledky porovnejte..., ...ve spolupráci s cílovými uživateli otestujte a porovnejte s dosavadním řešením..., ...experiment, který by měl validovat...

Vše, o čem bude řeč, souvisí s metodickou stránkou výzkumu – tedy kupříkladu i Vašich diplomových prací. Dílčí části pak například se statickým vyhodnocením dat.

Co vlastně lze zjišťovat či měřit a následně analyzovat (v rámci testování použitelnosti)?

Kognitivní kartografie zkoumá percepci map (i) za účelem zvýšení efektivity a jejich přizpůsobení potřebám konkrétních uživatelů, což je označováno jako studie použitelnosti.

Elzaker (2004) rozlišuje dva základní typy testování v kartografii – funkční („jak dobře mapa plní svůj účel“) a kognitivní („jak“ a „proč“ mapy fungují).

Tři základní ukazatele:

- effectiveness – správnost/chybovost,
- efficiency – rychlost,
- satisfaction – spokojenost, uspokojení.

Někdy se uvádí a používají také snadnost naučení (lerneability), zapamatovatelnost (memorability), přístupnost (accessibility, jestli a jak je uzpůsobeno pro znevýhodněné uživatele, např. s poruchami barvocitu), užitečnost (usefulness, jak dobře splňuje daný účel – špatně se měří), ...

Nejčastější jsou ale tři charakteristiky uvedené v odrážkách, lze je i vzájemně kombinovat (ale není to časté, většinou se hodnotí odděleně). Více o jejich kombinaci viz <https://usabilitygeek.com/usability-metrics-a-guide-to-quantify-system-usability/>

Někdy se výzkum zaměřuje také na způsob, jakým uživatelé řeší zadanou úlohu. Tedy na jejich strategii.

Jaké metody pro testování a hodnocení použitelnosti znáte?

a) Uživatelské testování:

• **Rozhovor**

- forma: většinou polostrukturovaný (existuje scénář, ale ne všechny otázky jsou známy předem)
- výhody: lze získat hluboký vhled do motivů, chování a názorů
- nevýhody: může být velká časová náročnost.
- klasifikace výzkumné metody: subjektivní a kvalitativní
- použití v kartografii je poměrně časté

• **Dotazník**

- forma: papírový nebo digitální (webový), může obsahovat otevřené i uzavřené otázky, Likertovy škály jsou rovněž časté
- výhody: umožňují rychlý a levný sběr dat od velkého množství respondentů,
- klasifikace výzkumné metody: kvantitativní a subjektivní,
- použití v kartografii je velmi časté, příklady: GoogleForm, LimeSurvey, SurveyMonkey nebo Survio

• **Eye-tracking**

- forma: sledování pohybu očí pomocí kombinace speciálního hardware a software (může být jak laboratorní/desktop, tak i mobilní – brýle)
- výhody: zkoumá se zrakové vnímání a mapy jsou vnímány primárně právě zrakem, umožňuje zkoumat i uživatelské strategie
- nevýhody: neřeší se periferní vidění, nastavení softwarového nástroje a vyhodnocení dat může být náročné, mobilní eye-tracking může být delší dobu nepohodlný
- klasifikace výzkumné metody: objektivní, kvalitativní i kvantitativní
- použití v kartografii je relativně časté (viz výhody)
- příklady: zařízení EyeTribe, SMI; software OGAMA.

• **Psychofyziologické měření**

- forma: zahrnuje řadu jednotlivých metod, např. EEG (elektroencefalogram), EDA (elektrodermální aktivita), EMG (elektromyografie), face-tracking či pupilometrie (měření změny velikosti zornice)
- nevýhody: většinou obtížná interpretace, obvykle drahý hardware
- klasifikace výzkumné metody: objektivní a kvantitativní
- v kartografii se používá zřídka, jsou popsány dílčí použití EEG nebo pupilometrie (souvisejí s eye-trackingem)

• **Think aloud (přemýšlení nahlas)**

- forma: metoda je vyvinutá k zachycení toku informací (myšlenek), o kterých testující subjekt přemýšlí, zatímco pracuje v testovaném uživatelském prostředí. Existuje i retrospektivní varianta (tzn. přemýšlení o ukončené práci)
- výhody: lze získat postoje a preference zároveň s testováním namísto pokládání otázek až po jeho skončení. Přemýšlení nahlas může participantům pomoci se lépe koncentrovat na zadaný úkol a zároveň si mohou lépe uvědomit, jak o úkolu přemýšlejí.
- nevýhody: pro někoho může být přemýšlení nahlas nepřírozené a může jej považovat za rušivý vliv. Přemýšlení nahlas zároveň zpomaluje proces myšlení.
- klasifikace výzkumné metody: subjektivní i objektivní, kvalitativní i kvantitativní (naměřená data = audiozáznam je možné analyzovat oběma způsoby)

- **Focus group (skupinové diskuse)**
 - forma: technika pro vedení skupinových rozhovorů
 - nevýhody: podmínkou je zkušený moderátor
 - klasifikace výzkumné metody: subjektivní a kvalitativní
 - v kartografii se používá spíše ojediněle
- **Třídění karet**
 - forma: uživatelé mají za úkol roztřídit pojmy či objekty do skupin a ty případně pojmenovat. Cílem testu je pochopit na základě čeho uživatelé třídí a prioritizují informace.
 - klasifikace výzkumné metody: subjektivní a kvalitativní i kvantitativní
 - v kartografii se používá zřídka, by bylo možné jej použít například na hodnocení mapových znaků (symbolů).
- **A/B testování**
 - forma: uživateli je náhodně přiřazena jedna ze dvou variant produktu a následně je měřen vliv varianty na jejich práci. V první fázi je nezbytné navrhnout dvě varianty produktu (např. webové mapy).
 - klasifikace výzkumné metody: objektivní a kvantitativní
 - v kartografii se používá, i když se to třeba explicitně neuvádí. V případě interaktivních vizualizací lze kombinovat s user loggingem.
- **User logging, screen logging, mouse logging**
 - forma: software, který zaznamenává, co je zobrazeno na obrazovce, pohybu myši, kliknutí, zmáčknuté klávesy apod.
 - výhody: vhodné pro analýzu webových stránek nebo interaktivních aplikací.
 - nevýhody: pro hodnocení statické mapy je tato metoda nepoužitelná
 - klasifikace výzkumné metody: objektivní, kvalitativní i kvantitativní
 - v kartografii se používá, ale spíše zřídka; existuje řada nástrojů pro web Inspectlet, Mouseflow, HotJar nebo český projekt mYx
- **Pozorování**
 - forma: zjišťuje se způsob interakce zkoumané osoby nebo skupiny osob s nějakým produktem. Mohou se stanovit scénáře nebo nemusí. Výzkumník si vede poznámky o tom, na jaké problémy zkoumaná osoba narazila, jak je vyřešila, a jaké funkce se musí u produktu upravit.
 - nevýhody: klade vysoké nároky na pozorovatele, může být časově náročné
 - klasifikace výzkumné metody: kvalitativní i kvantitativní (když je strukturováno)
 - v kartografii má smysl u interaktivních vizualizací
- **Mixed-research design** je označení pro kombinaci více výzkumných metod, a to jak kvalitativních, tak kvantitativních.

b) Expertní hodnocení:

- **Heuristické hodnocení**
 - forma: expertní hodnocení, které zkoumá rozhraní a posuzuje jeho soulad s principy použitelnosti (heuristikami).
 - výhoda: nevyžaduje uživatele, může být realizováno v krátkém čase
 - nevýhoda: může být subjektivní (ovlivněné znalostmi posuzovatele)
 - v rámci kartografie se použilo například na (geo)portály veřejné správy, př.:
<https://dk.upce.cz/handle/10195/37472>

- **Kontrola funkčních prvků (feature inspection)**
 - forma: využívá pro hodnocení seznam funkčních prvků potřebných pro dosažení typických úkolů (výhody a nevýhody viz heuristické hodnocení)
- **Kontrola konzistence (consistency inspection)**
 - forma: hodnocení zkušených designérů, kteří porovnávají své znalosti se návrhem zkoumaného produktu
- **Kontrola standardů (standart inspection)**
 - forma: kontrola vůči příslušným normám.
- **Kognitivní průchod (cognitive walkthrough)**
 - forma: jeden nebo více expertů vypracovávají řadu úkolů a otázek z pohledu uživatele. Cíl je získat a pochopit první zkušenosti nových uživatelů. Kognitivní průchod se aplikuje u složitějších systémů a nástrojů pro vývoj softwaru.
- **Pluralistický průchod (pluralistic walkthrough)**
 - forma: skupinovém hodnocení produktu, skupina expertů představuje role reálného uživatele a společně prochází aplikací s cílem vzájemně diskutovat o nalezených nedostatcích.
- ...

Lze výzkumné metody nějak rozčlenit či vzájemně porovnávat?

- Expertní hodnocení a uživatelské testování
- subjektivní (attitudinal; co uživatel říká) a objektivní (behavioral; co uživatel dělám),
- kvalitativní (pozoruju, odpovídám na otázky „proč“ nebo „jak“) a kvantitativní (měřím, otázky „kolik“),
- dle kontextu použití (v laboratoři, v přirozeném prostředí – doma, v terénu, ...).

Jak ovlivňuje uživatelské testování současná situace kolem COVIDu?

- **Testování na dálku (remote testing)**
 - **synchronní** – testování s moderátorem (př. Skype, MS Teams, Mikogo.com)
 - **asynchronní** – testování bez moderátora
 - výhody: možnost oslovit více uživatelů, šetří čas všem svým účastníkům, testování probíhá většinou v přirozeném prostředí
 - nevýhody: nižší stupeň kontroly (nevyzpytatelnost) větší požadavky na techniku na straně uživatelů (na její konzistenci)
 - v kartografii se moc nerozvinulo (zatím)
 - Lze nevýhody nějak řešit?

Jak správně popsat (dokumentovat, prezentovat) testování použitelnosti?

- Mám vymyšlené/zadané výzkumná otázky/hypotézy?
 - Ano mám → uvedu je, viz dále.
 - Výzkumná otázka je otázka, např.: *Liší se muži a ženy v aktivitě při editaci dat OSM? Jak ovlivňuje míra detailu 3D modelu města rychlost orientaci uživatelů v něm?*

- Hypotéza je tvrzení, které se snažíte potvrdit (nebo vyvrátit): *Ženy jsou aktivnější při editaci OSM. Uživatelé se rychleji zorientují ve zjednodušeném 3D modelu města.*
 - Ne, nemám → je to exploratorní výzkum. Tuto informaci uvedu. *Např. když se snažíte identifikovat chyby při ovládání mobilní aplikace.*

- Co bych měl(a) napsat o účastnících testování:
 - Počet
 - Věkovou a genderovou strukturu
 - Vzdělání (obor, ročník) – někdy stačí geo-/negeo-, či obecně obor studia (technický, sociální, přírodovědný, ...), pokud řešíme například problematiku vzdělávání tak je dobré popsat co nejpřesněji.
 - Předchozí zkušenosti – např. práce s mapou, práce s počítačem, chytrým telefonem – záleží co zkoumáme.
 - Zrakové vady, lateralitu (pravák versus levák – pokud řeším interakci), zdravotní obtíže (např. epilepsie – neměl by pracovat s VR/HMD)
 - Pokud některé z vyhodnocování vyloučím, uvést kolik a proč (např. nepochopili zadání, nedokončili test, měli zrakovou vadu, ...)

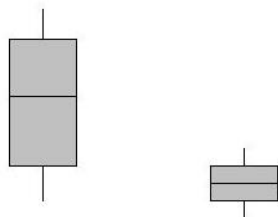
- Co uvádět při popisu průběhu testování
 - Kdy (datum či rozmezí) probíhalo
 - Jaké výzkumné metody byly použity – *příklady viz výše.*
 - V jakých podmínkách, např. *laboratoř VGE na GÚ, v učebně Z7. V učebnách Z1 a Z7, při zachování stejných světelných podmínek a rozlišení monitorů, on-line, ...*
 - Konkrétní design
 - Within-subject = vnitrosubjektový plán experimentu, př.: uživatel pracuje se všemi variantami, *např mapy = jak ortofoto, tak topografická mapa.*
 - nevýhoda: efekt učení
 - Between-subject = mezisubjektový plán experimentu, př.: *uživatel pracuje pouze s jednou variantou zkoumané mapy (jen ortofoto, nebo jen topografická mapa)*
 - nevýhoda: inter-individuální rozdíly
 - nebo nějaké jejich kombinace či multivarianční experiment
 - Schéma experimentu (obrázek), dále popsat:
 - Rozdělení do skupin (náhodné, na základě nějaké konkrétní charakteristiky – gender, věk, zkušenosti)
 - znáhodňování a vyvažování pořadí úloh (counterbalancing) – ano/ne, pokud ano, tak kompletní nebo nekompletní
 - Popsat s jakými stimuly uživatelé pracovali, včetně ukázek (obrázků).
 - Popsat jakou úlohu/úloh řešili, např. *Uživatelé měli kliknout na místo v terénu, které se podle nich nachází v nejvyšší nadmořské výšce; nebo: Uživatelé měli za úkol kliknutím označit variantu mapového znaku, která se jim nejvíce líbila.*

Nasbíral jsem data. Co teď mám dělat s tou statistikou?

- Mám data, např.: rychlost hledání zadaného symbolu na podkladě topografické mapy a ortofotomapy; data o výnosu naměřená a výsledek prediktivního modelování výnosu; data povrchové teplotě a NDVI; ...
- Exploruju data = popisná statistika, grafy: histogramy, box-ploty
- Zjistím, jestli zpracovávané hodnoty mají normální rozdělení.
 - o Pokračuju v exploraci dat, podívám se na histogram
 - o Otestuju normalitu – např. Shapiro-Wilk test (jeho výhodou že funguje i na menší vzorky – např. nižší desítky uživatelů). Když je p hodnota (p-value) větší než α (často se používá 0,05 = 95% hladina významnosti), tak je rozdělení normální.
 - o Když je rozdělení normální používáme parametrické metody. Pokud rozdělení není normální, používáme neparametrické metody (což je často v okamžiku, když pracujeme s menšími vzorky). Když to lze použiju parametrické metody (jsou „lepší“, „silnější“, „čistší“)

		Parametrické metody	Neparametrické metody
Popisná statistika (tabulky, box-ploty, hodnoty v textu)		Průměr, směrodatná odchylka a její násobky	Medián, mezikvartilové rozpětí
Jeden vzorek a celá populace	Testy středních hodnot	Jednovýběrový t-test	Jednovýběrový Wilcoxonův test, znaménkový (signed) test
Dva vzorky – nepárové		Nepárový t-test	Mann-Whitney test = Wilcoxon rank-sum test
Dva vzorky – párové		Párový t-test	Wilcoxon signed-rank test, znaménkový (signed) test
Tři a více vzorků – nepárové		ANOVA	Kruskall-Wallisův test = neparametrická ANOVA
Dva vzorky – nepárové	Test rozptylu	F-test	Levenův test
Korelace (většinou dva vzorky)		Pearsonův koeficient	Spearmanův koeficient

- o Párový versus nepárový test:
 - Párový test př.: *jeden uživatel pracoval jak s ortofotomapou, tak topografickou mapou.*
 - Nepárový test př.: *S ortofotomapou pracovalo 10 uživatelů, deset jiných pracovalo s topografickou mapou*
- o Test středních hodnot versus test rozptylu:



- Záleží, co používám za statistický software (Excel, Statistika, R, ...), ale např. Statistika interpretaci výsledků testu usnadní (statisticky významné záležitosti jsou červeně).

Co mám pak psát v diskusi?

- Srovnání zjištěných výsledků s jinými podobně zaměřenými studii
- Výhody a limity (slabiny) vlastního výzkumu – kriticky zhodnotit informace o průběhu experimentu, jak se do budoucna vyrovnat s limity výzkumu.
- S tím souvisí ekologická validita (jak se blíží „realitě“, co bylo zjednodušeno)
- Jak by se na provedený výzkum dalo navázat, na co se zaměřit v budoucnu