

PŘÍLOHA 1 K DIPLOMOVÉ PRÁCI T. VANÍČKA (2022)

„VYUŽITÍ METODY THINK-ALOUD V KARTOGRAFICKÉM VÝZKUMU“

THINK-ALOUD

**PRAKTICKÝ PRŮVODCE PRO TVORBU
SOUBĚŽNÝCH THINK-ALOUD EXPERIMENTŮ
NEJEN V KARTOGRAFICKÉM VÝZKUMU**



Tomáš **VANÍČEK**

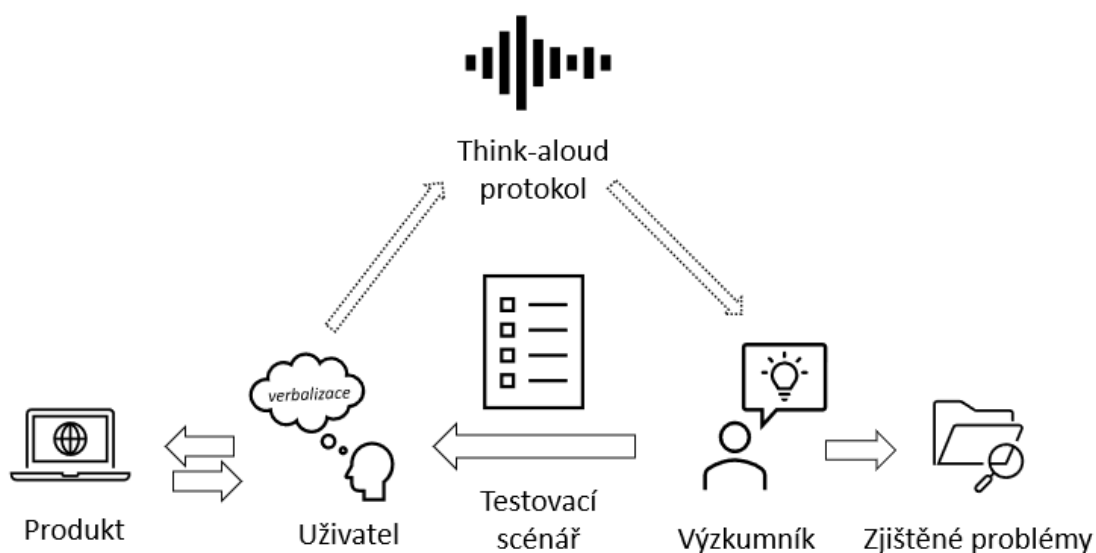
Olomouc 2022

Vítejte u stručného **praktického průvodce pro tvorbu a hodnocení souběžných think-aloud experimentů** (tzv. CTA). Tento průvodce byl vytvořen v rámci diplomové práce „Využití metody think-aloud v kartografickém výzkumu“ jako její příloha. Zároveň je k dispozici online ke stažení na webových stránkách diplomové práce.

Průvodce přináší kontinuální postup od přípravy až po vyhodnocení think-aloud experimentů. Jednotlivé kroky jsou zvoleny jednak na základě zkušeností autora sesbíraných během tvorby diplomové práce, ale také podle široké řady odborné literatury a jiných studií věnovaných dané metodě. Jednotlivé kroky jsou charakterizovány stručně a poskytnuty jsou jen základní informace. V případě zájmu o odbornější a hlubší výklad jsou analogické postupy součástí zmíněné diplomové práce autora.

Co je to think-aloud metoda?

Think-aloud je metoda používaná pro shromažďování informací při testování **použitelnosti produktu**. Využívá se ale také v psychologii a řadě dalších společenských věd. **Účastníci** testování v kontrolovaném prostředí **pracují s produktem** a plní specifickou připravenou sadu otázek a úkolů, během které **přemýšlí nahlas**. Respondent **nahlas verbalizuje** veškeré své **myšlenky** průběžně po celou dobu testování. Řeč může obsahovat řadu myšlenek participantů, např. na co se dívají, na co myslí, co dělají nebo jak se cítí. Řeč uživatele je zároveň **zaznamenávána do záznamového zařízení** pro archivaci komunikace a následnou analýzu. Tento princip umožňuje pozorovateli hlouběji nahlédnout do kognitivních procesů účastníka. Audio nebo video záznam verbalizací je poté transkribován a analyzován jak **kvalitativně**, tak **kvantitativně**. Vzniká tak hodnocení použitelnosti produktu. Pozorovatel si také vytváří poznámky během procesu testování o tom, co účastníci říkají a dělají, aniž by zasahoval do průběhu procesu think-aloud experimentu.



Princip think-aloud metody

Krok 1: Výběr (kartografického) produktu

V první řadě je nutné zvolit **produkt**, který bude předmětem testování a u kterého chce výzkumník zjistit jeho použitelnost. Může se jednat o libovolný **fyzický produkt** (např. mapa, atlas, model), ale i **systémové rozhraní** (např. webová stránka, webová mapová aplikace, software). V případě komplexnějšího produktu (jako je například atlas), je nutné se zaměřit jen na dílčí části (témata, prvky), které budou podrobeny testování. Vhodné je kombinovat think-aloud s eye-tracking testováním.

Krok 2: Tvorba testovacího scénáře

Před zahájením experimentu je potřeba vytvořit **testovací scénář**. Jedná se o otázky a úkoly pro respondenty. Vytvořené úlohy by měly být v souladu s tím, co je předmětem testování (např. kapitola v atlase, kompoziční prvky, uživatelské rozhraní apod.). Některé úlohy však nejsou vhodné, protože zahrnují buď „příliš verbální“ nebo neverbální komunikaci. Neverbální informací může být například přečtení nadpisu. Uživatel sice přečte nadpis, ale pro výzkum tato informace nemá přínos. Výsledky jsou tedy velmi ovlivněny položenými otázkami. U verbální komunikace nejsou některé úlohy vhodné z hlediska jejich povahy, tj. například řízení letového provozu nebo psychoterapie. Pro tyto oblasti je zkrátka nemožné vytvořit vhodné úlohy, které by naskytly pohled do uživatelského myšlení. Také může být náročnější vytvářet úkoly v oblastech, které jsou svou povahou komplexnější, například řešení fyzikálních problémů nebo architektonické návrhy. Podle Somerena (1994) je potřeba provést následující úvahy:

- 1) Je úkol na takové úrovni obtížnosti vzhledem k patřičné kognitivní zátěži na daného uživatele?

Respondenti by měli být schopni úlohu řešit automatizovaně, ale zároveň by měla být úloha dostatečně obtížná.

- 2) Je úloha reprezentativní pro řešený problém a je zároveň možné ji chtít vyřešit po uživateli?

Výzkumník je při tvorbě testovacího občas v pokušení zvolit takové otázky, které jsou neobvyklé a obtížné, např. příliš detailní. Rizikem však je, že výsledkem budou údaje, které jsou pro studovaný problém relevantní jen okrajově, protože si úkol respondent „ohne podle sebe“. Před zahájením experimentu je vhodné ukázat testované otázky jinému odborníkovi, který pomůže s jejich finalizací tak, aby byly obtížné a zároveň relevantní. Po skončení experimentu je také vhodné se zeptat participantů, jak se při testování cítili a zda pro ně byly otázky srozumitelné nebo v jistém smyslu neobvyklé. Průběh testování by měl ideálně trvat krátce, proto lze metodu uplatňovat pouze u malého souboru problému. Například u atlasu je dobré zvolit pouze vybrané okruhy. Nejlepším způsobem, jak efektivně využít metodu v kombinaci s časovým omezením je pravděpodobně kombinovat metodu s jinými metodami (např. eye-tracking), které umožňují získat obraz o obecnosti výsledků. Účelem metody je získat údaje o kognitivním procesu účastníka a o jeho postojích na produkt. Situace by ve výsledku měla být taková, aby otázky odpovídaly optimální formě složitosti a průběh řešení úkoly byl verbalizovatelný z pracovní paměti. Je také vhodné volit úkoly otevřenějšího charakteru, aby se během jejich řešení mohl uživatel rozpovídat. Dalším důležitým faktorem pro vhodné otázky je předmět testování a jeho cíl. Je nutné dopředu vědět, jaký produkt je testován, co tento produkt umí nebo měl by umět a co je od testování očekáváno. V případě webové mapy

totiž mohou být shromažďována primární data jako počet kliknutí myši během vykonávání úkolu, čas strávený při plnění úkolu anebo také úspěšnost vykonaných úkolů. Je tedy nutné si před tvorbou otázek určit, co je předmětem testování a co má být jeho cílem (kvalitativní nebo kvantitativní charakteristiky).

Testovací scénář je dobré vytisknout na papír a vložit do fólie, aby byl respondentovi vždy na očích. Zároveň je dobré mít otázky pod sebou s větším rozstupem anebo barevně od sebe odlišit, aby se uživatel při návratu k otázkám neztrácel. Součástí by měla být i tzv. „rozehřívací úloha“. Tato logická úloha by měla být představena respondentovi ještě před začátkem samotného testování. Základní účelem úvodní úlohy je to, aby si participant zvykl na přemýšlení nahlas a zároveň si pozorovatel zjistil, „jak na tom uživatel s verbalizací je“. Zadání může vypadat například následovně:

„V 6 hodin ráno byla teplota 19,4 °C. Mezi 6:00 a polednem teplota stoupla o 3,8 °C. Mezi polednem a osmnáctou hodinou teplota klesla o 2,5 °C. Jaká byla teplota v 18 hodin?“

Jedná se o velmi jednoduchou úlohu. Pokud ji uživatel dokáže nahlas vyřešit, může být přistoupeno k samotnému testování.

Krok 3: Výběr reprezentativních uživatelů

Pro získání kvalitních dat je nutné zvolit reprezentativní a vhodné uživatele pro výzkum. S tím souvisí i **počet uživatelů** k testování. Důležité je určení **cílové skupiny produktu**. Právě potenciální uživatelé produktu by měli být součástí testování. Výzkum, potažmo produkt, může být zaměřen jen na určitý druh osob. Cílem výzkumníka je získat náhodný vzorek těchto osob pro testování a výsledky zobecnit na všechny osoby tohoto druhu. Také je často obtížné získat přístup k odborníkovi. Dvěma důležitými vlastnostmi subjektů jsou **stupeň odbornosti** a **schopnost verbalizace**. Pro výběr respondentů může být uplatňována i tzv. metoda sněhové koule, kdy noví testující jsou nominováni osobami, které již jsou ve výborovém souboru. Je však lepší upřednostňovat spíše participanty s extrovertním chováním.

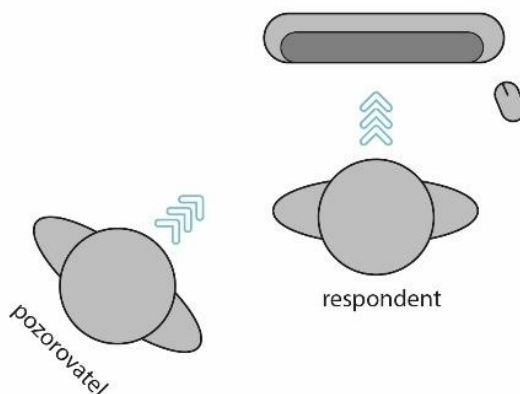
Jednotlivec nedokáže zjistit všechny problémy použitelnosti produktu. Z tohoto důvodu je nutné zvolit optimální počet respondentů tak, aby bylo ideálně dosaženo co největšího pokrytí všech problémů produktu. Podle Nielsena (2012) existuje jednoduchá odpověď pro počet respondentů při kvalitativních experimentech, a to je pět. Testování s pěti participanty totiž umožní najít téměř stejný počet problémů s použitelností produktu (až 80 %) jak s mnohem větším počtem účastníků testu. Navíc nezáleží, jaký produkt je testován (aplikace, webové stránky, výrobek, mapa...), s pěti respondenty se lze téměř vždy přiblížit k maximální efektivitě, která je určována jako poměry mezi počtem zjištění, nákladů na výzkum a času (Nielsen, 2012). Problematice o počtech uživatelů pro testování se zabývali i další výzkumníci. Podle Stefana a kol. (2009) je neoptimálnější počet testujících pro metodu think-aloud **šest respondentů**. Tento počet testujících v jejich výzkumu odhalil až 84 % problémů použitelnosti. Skupina Nielsena a Normana porovnála počet zjištění s počtem uživatelů všech 83 jejich studií použitelnosti. V rámci těchto mnoha projektů testování více uživatelů nevedlo ke zdatelně větším zjištěním (Nielsen, 2012). Při metodě think-aloud však může dojít k situaci, kdy testovaný uživatel plně neverbalizuje své myšlenkové pochody. V tomto případě je test buď zahrnut do hodnocení použitelnosti produktu anebo zahrnut s tím, že výpověď konkrétního respondenta není úplná (např. nebylo vypátráno nových zjištění). Pokud nastane tato situace a výzkumník si je vědom vzniklého stavu, pak je vhodné pozvat k testování další subjekt ze stejné cílové skupiny. Tím se však navyšuje požadavek pro vyšší počet testovaných, ale i náklady na výzkum.

Odborníci na daný produkt (např. mapa ~ kartografové) mohou patřit mezi skupinu testovaných. Získají se tak jejich expertní názory. Na druhou stranu může dojít k několika problémům. Odborníci jsou schopni velice dobře provést určitý úkol, ale nedokážou vysvětlit, jak přišli na správnou odpověď („prostě jsem viděl, že to musí být tohle“). Jejich znalosti (a ideje na produkt) tak nejsou plně verbalizovány. Z praktického hlediska však experti dokážou své znalosti explicitně vyjádřit v následné diskusi, kterou je vhodné provést a zaznamenat. Experti mohou být ohledně svých odborných znalostí tajnůstkářští a mohou se zdráhat nahlédnout někomu jinému do svého skutečného chování při řešení problémů. Většina expertů si je dobře vědoma toho, že své odpovědi nemohou snadno zdůvodnit (jednají rutinně), ale nemusí chtít, aby to jejich posluchači věděli. Z tohoto důvodu mohou zaujmout racionálnější styl zdůvodňování. V praxi však může expert produkovat racionalizace, které nemají žádný vztah ke skutečným odborným znalostem, tj. skryje nejasné, špatně pochopené oblasti své odbornosti, a to buď ve snaze pomoci pozorovateli tím, že věci zjednoduší a vyhne se nepřehledným detailům, nebo tím, že výzkumníkovi pomůže najít zdůvodnění odpovědi. Expert by proto měl být poučen, že je důležitější, aby byly výpovědi přirozené a přímé, než aby byl srozumitelný osobě, která data pořizuje. Nesrozumitelné části lze vždy dodatečně vyjasnit, ale chybějící myšlenky a znalosti nelze obnovit. To znamená, že je třeba dbát na to, aby byla zajištěna ochota experta spolupracovat na sezeních, kde se přemýšlí nahlas.

Krok 4: Příprava prostředí a techniky

Samotné testování je pravděpodobně tou nejjednodušší částí z pohledu výzkumníka, protože jsou již vybráni respondenti a je připraven **testovací scénář**. Na druhou stranu i drobné chyby v postupu experimentu mohou způsobit, že získané údaje budou téměř nepoužitelné. Ještě před pozváním respondenta do místnosti je nutné připravit techniku a prostředí, následně uvést respondenta a představit mu, proč tu je a co je jeho hlavním úkolem. Výstupem úspěšného testování by měla být nahraná data s výpověďmi účastníka (tzv. raw protocol).

Hlavním požadavkem pro úspěšné provedení metody je **kontrolované prostředí**, ve kterém probíhá testování. Pozorovatel se musí ujistit, aby se účastník cítil v místnosti dobře. Základem je **tichá** a ideálně **odhlučněná** místnost. Pro testovaného je dobré mít připravenou **pohodlnou židli** a **sklenici vody**, zejména v případě, pokud má experiment trvat delší dobu. Je také dobré mít na paměti, že výzkumník by měl co nejméně (ideálně vůbec) zasahovat do průběhu experimentu, aby neovlivnil myšlenkové procesy účastníka. Z tohoto důvodu by měl být pozorovatel vzdálen od uživatele.



Poloha pozorovatele vzhledem k testovanému uživateli

Základní faktorem je mít připravený **kvalitní mikrofon** a ujistit se o jeho funkčnosti. Před uvedením respondenta je nutné vždy otestovat, zda je mikrofon funkční a zvuk je úspěšně nahráván do počítače. V případě kombinace s eye-tracking testováním také otestovat funkčnost ostatních zařízení (eye-tracker, kamera atd.). Kromě kontroly technických zařízení je dobré si připravit i standartní výbavu výzkumníka jako papír na poznámky, psací potřeby a testovací scénář. Mimoto nachystat i drobné (sladké) odměny pro respondenty.

Krok 5: Uvítání respondenta

Jakmile je prostředí nachystáno a byla otestována veškerá zařízení, může dojít k uvítání respondenta na testování. Základem je **úvodní seznámení** a **poděkování** participantovi za příchod. Po uvedení je potřeba účastníka seznámit se **smyslem testování** a jeho **hlavními cíli**. Pokyny u této metody jsou jednoduché, v podstatě stačí říct: „proved' úkol a nahlas říkej vše, co tě napadne.“ Vhodné je instrukce pro respondenta sepsat dopředu na papír a ty mu přečíst. Například Hamel (1990) ve své studii architektonického designu podával následující úvodní informace:

„Za chvíli dostaneš designový úkol. Prosím, abys tento úkol vykonal/a způsobem, jakým jsi zvyklý/á ve své každodenní praxi. Je důležité, abys říkal/a nahlas vše, co si myslíš nebo děláš během úkolu.“ Podle Somerena (1994) není dobré používat fráze jako: „Řekni mi, co si myslíš.“ Lidé si mohou myslet, že se jich výzkumník ptá na jejich konkrétní názor nebo na hodnocení jejich myšlenek. Zároveň by tento typ fráze naznačoval, že by participant nad problémem musel skutečně přemýšlet a necítil by se tedy pohodlně. Instrukce by také měly být striktní a rychlé. Čím více pokynů je řečeno, tím více si budou subjekty vytvářet vlastní interpretace toho, co se po nich chce. Také je vhodné se zmínit o **ochraně údajů** a pro jaké účely budou data využita. Důležité je vysvětlit, že s údaji bude nakládáno přísně důvěrně. Ochrana soukromí je nejen otázkou etiky a právní praxe, ale je důležitá i pro samotný výzkum. Kromě toho může být vhodné zdůraznit, že smyslem je zjistit, jakým způsobem lidé řeší problémy a že nejde o zjišťování jiných nevědomých emocí nebo skrytých myšlenek (jinak někteří participantů mohou mít výhrady k výzkumu). Důležitou informací pro participanta je, že **„každá odpověď je dobrá“** a že se nemusí bát, že by se **ztrapnil**. Pokud je subjekt velmi **nervózní**, může mu to bránit mluvit nahlas. To není zanedbatelný faktor. V tom případě je vhodné zmínit, že se po něm chce pouze říkat nahlas to, co „se mu honí v hlavě“ při práci s produktem a že jde jen o to zjistit, jakým způsobem řeší problém. Nervozita však může být problém jak pro testovaného, tak i pro výzkumníka, protože data nemusí být úplná. Tento problém nastává zejména u studentů. Vytvořit **atmosféru důvěry a lehkosti** je tedy nesmírně **důležité**. Poté lze účastníka informovat, aby se řídil **testovacím scénářem** a také že pozorovatel **nesmí do průběhu testování zasahovat**. Ještě před samotným začátkem testování je ale uživateli nutné zadat první **rozehřívací úkol**, který zajistí, aby se mozek naladil na proces **verbalizace**. Protože většina subjektů není zvyklá při řešení problémů přemýšlet nahlas, je vhodné před testováním nechat účastníka vyřešit nahlas určitý logický problém. Řešení tohoto úkolu nemusí být nahráváno. Při samotném testování by tak měl účastník snáze verbalizovat své myšlenky.

Krok 6: Zahájení testování

Pozorovatel by se měl nacházet **za respondentem**. Díky tomu dotazovaný netuší, jak se pozorovatel tváří. Zároveň by měl být po celou dobu pozorovatel v klidu, nehýbat se, nemluvit, nevydávat jiné zvuky a při psaní poznámek nedělat rámus. Pokud respondent řekne něco vtipného, neměl by na to výzkumník reagovat. Situace by měla být zaměřena na úkol a experimentátor by měl co nejméně zasahovat do myšlenkového procesu, aby neovlivňoval jeho průběh. Role experimentátora by měla být tedy **více zdrženlivá**. K zasahování by mělo docházet pouze tehdy, když subjekt přestane mluvit. Pak by měl pozorovatel subjekt pobídnout jen a pouze slovy: „Pokračuj v mluvení“. To obvykle stačí k tomu, aby subjekt nějakou dobu přemýšlel nahlas. Pro experimentátora průběh řešení nemusí být lehká práce a musí se ovládat, zvláště v případě, kdy má tendenci respondenta opravovat, když dělá chyby, nebo mu pomáhat, když se zasekne. Tomu je skutečně třeba se vyhnout. Experimentátor by tedy měl mít určité zkušenosti se shromažďováním údajů o přemýšlení nahlas a zejména by se měl vyvarovat zbytečných zásahů. Testování je obvykle nahráváno na diktafon nebo pomocí mikrofону zabudovaného k PC, přičemž audio nebo video soubor je uložen do počítače. Platí pravidlo, že před zahájením experimentu je vždy nutné **zkontrolovat funkčnost** nahrávacích nástrojů.

Krok 7: Konec testování

Jakmile je testování u konce, nemělo by být rozloučení s respondentem příliš rychlé. Bez participantů na testování by nedošlo k žádnému testování. Proto je potřeba si jich vážit. Důležitost účasti na testování byla možná zmíněna ještě předtím, než byl uživatel přizvat na experiment, přesto je nejdůležitější součástí u konce experimentu **velké poděkování** participantovi ze strany výzkumníka o účasti na tomto výzkumu. Je vhodné také zmínit, že jeho výpovědi mají hodnotu a že účastí na testování pomohl vývoj produktu ovlivnit. Kromě toho může mít uživatel obavy o tom, že jeho výkon nebyl ideální a že si je vědom chyb. V tomto případě je vhodné jej **informovat a uklidnit**, že nešlo o správnost odpovědí a ani o jeho výkon a že si vedl dobře, ať už byl průběh testování jakýkoliv. Zároveň je vhodné si vyžádat **kontakt**, pokud bude respondent souhlasit a informovat ho o výsledcích po ukončení studie. Zároveň může kontakt sloužit výzkumníkovi pro případná další testování. Konec testování je také poslední příležitostí odpovědět na jakékoliv otázky, které může respondent mít o testu nebo produktu. Jako hmatatelné poděkování také není od věci nabídnout respondentovi **odměnu**, například formou malé čokolády, což ocení převážně mladší participantí.

Někdy však po skončení testování přichází na řadu dotazník, který ověřuje aspekt spokojenosti uživatele při práci s produktem nebo testováním. Existují dva druhy dotazníků, které se běžně využívají, a to SEQ a SUS.

SEQ (Single Ease Question)

Základním principem SEQ je položení jediné otázky respondentovi po proběhlém testování. Většinou se jedná o otázku typu „jak obtížné bylo pro tebe dokončit tento úkol?“ Většina metrik použitelnosti založených na úkolech se zaměřuje na hledání objektivních parametrů, SEQ využívá podstatu uživatelského zážitku: jeho subjektivitu. Možná, že uživateli trvalo splnění úkolu déle, ale neměl takový dojem. Co když uživatel jen pomaleji reaguje? Nebo byl na chvíli roztěkaný? Uživatelské subjektivní hodnocení obtížnosti není o nic méně důležité než počet chyb, kterých se dopustil. Pokud například uživatelé

ohodnotili obtížnost úkolu v průměru známkou 4,8 (čím vyšší číslo, tím jednodušší úkol), je nutné se ujistit, jestli výsledky nejsou nižší (obtížnější) než je tato hodnota.

SUS (Systém Usability Scale)

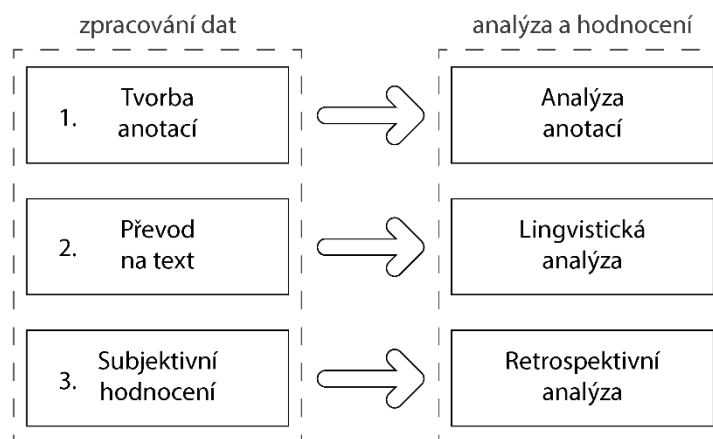
Pro ty, kteří nedůvěřují řešení s jednou otázkou, existuje seznam 10 otázek, známý jako „Stupnice použitelnosti systému“ (SUS). Na základě odpovědí získá produkt skóre na stupnici od 0 do 100 (každá otázka má hodnotu 10 bodů). Odpovědi jsou děleny podle stupnice od 1 do 5 (silně nesouhlasím nebo silně souhlasím). Seznam těchto otázek vypadá následovně:

1. Myslím, že bych tento produkt běžně využíval.
2. Myslím, že je tento produkt zbytečně složitý.
3. Myslím, že produkt je snadno použitelný (nebo vhodný pro zjištění informace).
4. Myslím, že bych potřeboval od někoho pomoc, abych byl schopen produkt používat.
5. Zjistil/a jsem, že různé funkce (nebo informace) produktu byly pochopitelné.
6. Myslím, že produkt obsahuje mnoho chyb.
7. Myslím, že spousta lidí by se naučila s tímto produktem pracovat velmi rychle.
8. Myslím, že produkt je velmi těžkopádný a špatný na používání.
9. Cítil/a jsem se pohodlně při práci s produktem.
10. Musel/a bych se naučit spoustu věcí, než bych mohl/a tento produkt používat.

Princip této stupnice se také hodí pro porovnání svého produktu s konkurenčním. Podle Panchenka (2021) průměrná hodnota SUS odpovídá 68 bodům. Výsledky nad 80 jsou považovány za vynikající.

Krok 8: Zpracování zaznamenaných dat

Výstupem úspěšného testování by měla být nahraná zvuková, případně i audiovizuální data s výpověďmi participanta (tzv. raw protocol). V tuto chvíli je pouze na výzkumníkovi, jak tato data využije, nicméně surová (raw) data pro komplexnější analýzu a zjištění použitelnosti produktu nepomohou a musejí být transkribována. To lze provést pomocí několika způsobů. Metodicky nejsprávnější variantou je segmentace záznamu a pro jednotlivé segmenty vypsání dané výpovědi. Další možností je převod řeči na text pomocí různých transkripčních nástrojů (nejvhodnější nástroje jsou uvedeny v diplomové práci) a poté provedení lingvistické analýzy nebo doplnění převedeného textu do anotací. Poslední možností je využití veškerých vzniklých materiálů a provedení subjektivní retrospektivní analýzy, ovšem tato metoda by měla sloužit pouze jako doplněk k předchozím možnostem.



Metody zpracování a analýzy dat

Nejvhodnější možností zpracování dat je segmentace záznamu a tvorba anotací. Separace získaných záznamů, vytváření segmentů a jejich anotování je nejdůležitější částí před analýzou. Poté, co jsou získány zvukové záznamy s výpověďmi, je nutné provést jejich **transkripci**. Transkripce obvykle znamená doslovný přepis do textové podoby, ovšem ne v případě metody think-aloud. Přesto v případě doslovného přepisu, což může být někdy náročné, lze aplikovat následnou lingvistickou analýzu. U metody think-aloud znamená transkripce **separaci záznamu na jednotlivé segmenty** a tyto nové segmenty je možné přepsat. V diplomové práci bylo využito programu ELAN pro segmentaci a anotaci záznamů (více je o něm uvedeno v textu diplomové práce). Přepisování záznamu je však časově náročné. Problémem u výpovědí může být srozumitelnost záznamu, protože lidé mohou mít ve zvyku nedokončovat slova, mohou mumlat, subjekt se může přerušit uprostřed slova, může něco zašeptat, může vydávat podivné zvuky, zkrátka ne vždy je možné rozumět všemu, co přesně bylo řečeno. Takovéto části záznamu by měly být označeny vlastní segmentem s anotací jako „nesrozumitelné“. Podle Somerena (1994) v psychologickém výzkumu je každá výpověď relevantní, a tak by měly být anotovány veškeré části záznamu, kde respondent nebo i pozorovatel mluvil, tzn. anotovat vše, co je v záznamu rozpoznatelné. Obecně platí, že hodnotitel by se měl snažit anotovat záznam co nejvěrněji a co nejvíce se držet toho, co dotyčný řekl. Dlouhé mlčení může být anotováno jako „ticho“. Veškeré anotace a jejich popis se **řídí kódovacím schématem**, které je výzkumníkem navrhováno před započítím transkripce záznamu. Pokud je testování nahráváno i na video, není od věci vytvořit anotace i pro činnosti respondenta, např. „subjekt bere tužku“ nebo „subjekt listuje v atlase“. Opět by však význam anotací měl být vysvětlen v kódovacím schématu. Někdy však není nutné anotovat celý záznam. Efektivnější může být kódovat pouze části, které jsou pro výzkumníka zajímavé, například přináší nové informace nebo pokud výzkumníka zajímá jen určitá aktivita (např. které prvky v mapě jsou využívány). Obecně lze tedy říct, že tvorba anotací je specificky dána vytvořenou výzkumnou otázkou, kterou chce pozorovatel zjistit, tzn. jaké aspekty produktu podléhají testování. Na základě těchto testovaných prvků a témat může být provedeno logování aktivity uživatele pomocí tvorby anotací, tj. přiřazení aktivity ke konkrétní části záznamu. Anotace jsou poté navrženy právě na základě zvolených otázek. Například pokud je předmětem testování posuzování legend u jednotlivých map, pak anotace mohou být detailněji děleny pro dané legendy v mapách. Tvorba anotací však podléhá vytvořenému kódovacímu schématu, které je nutno nejprve navrhnout.

Naměřené záznamy jsou nestructuralizovaným zdrojem informací a musejí být zakódovány podle kódovacího schématu. Kódovací schéma si tvoří sám výzkumník,

případně více výzkumníků u společné práce po jejich vzájemné dohodě, a podle něj jsou následně zakódovány všechny nahrávky k dané studii. Kategorie pro klasifikaci částí záznamu výzkumník volí podle výzkumné otázky a detailnosti strukturalizace záznamu. Pro podrobnější klasifikaci záznamu je vhodné zvukový záznam kombinovat s dodatečným exaktnějším záznamem, například videem s eye-tracking záznamem. Právě **kombinace zvukového záznamu s eye-tracking záznamem** může být velmi přínosná a může přinést exaktnější hodnocení.

Kódovací schéma použité v diplomové práci pro hodnocení atlasu

| Kategorie | Zkratka | Popis | Příklad |
|---|---------|--|--|
| Čtení otázky/úkolů | O | Uživatel čte otázku nebo její části | „V atlase naleznete téma vodstvo a popište mapy v této části.“ „...téma vodstvo a popsat...“ |
| Orientace v atlase, pročítání obsahu | A | Uživatel listuje atlasem, hledá patřičná témata, pročítá obsah | „...vracím se zpátky na obsah, téma doprava, jedu zase svrchu. Doprava hledám. Doprava je to strana 127, přecházím na stranu 127.“ |
| Čtení popisu mapy, textu nebo nadpisu | P | Uživatel čte popis v okolí mapy | „Městská hromadná doprava je v kraji provozována ve 13 městech.“ |
| Orientace v mapě a legendě | M | Uživatel hledá informaci a orientuje se v dané tematické mapě | „...úzkorozchodná regionální trať je značená čárkovanou a ta je v mapě mezi Osoblahama a Třemešnou...“ |
| Čtení grafických a infografických prvků | G | Uživatel hledá informaci v doplňujících infografikách, grafech a obrázcích | „...na to je tady infografika s tím, že je to přímo rozebrané na auto cyklo autobusy a nákladní automobily...“ |
| Nalezení informace a odpovědi | I | Uživatel našel požadovanou informaci | „...takže to nejspíše bude Odra, už jenom z čisté úvahy dle mapy.“ |
| Jiné | J | Jiné poznámky a úkony uživatele při přemýšlení nahlas a uživatelův názor | „Možná bude dobrý se podívat do seznamu, jestli tam něco bude.“ „Tak to vidím, že jsem se spletl...“ |

Krok 9: Analýza a vyhodnocení produktu

Jakmile jsou data zpracována podle jedné z metod zpracování dat, je možné přistoupit k jejich analýze. V předchozím kroku bylo provedeno anotování záznamu, přesto existuje více variací analýz nad získanými daty. Nejvhodnější metodou je analýza anotací pomocí jednoduchých anotačních nástrojů pro zvuková data (např. ELAN). Tato metoda by měla být primárním výběrem. Další možností je použití komplexnějších nástrojů pro lingvistickou analýzu textových dat (např. pomocí Sketch Engine). Doplňující alternativou je subjektivní hodnocení záznamu na základě empirií výzkumníka pomocí retrospektivní analýzy s využitím veškerých vzniklých materiálů.

Pro posouzení použitelnosti určitých částí atlasu může být využito základních metrik použitelnosti. Podle Paschenka (2021) jsou **metriky použitelnosti** systémem měření

- efektivity** (effectiveness) – např. úspěšnost a přesnost odpovědi,
- účinnosti** (efficiency) – např. čas potřebný k provedení úkoly,
- a **spokojenosti** (satisfaction) – postoj a názor uživatele,

uživatelů při práci s produktem. Tyto metriky se často používají v kartografickém výzkumu efektivity map. Pomocí strukturovaných zvukových záznamů lze tyto metriky snadno aplikovat. Na druhou stranu zde dochází k upozadění myšlenky kvalitativního testování a do popředí vystupují kvantitativní údaje. Ve většině případů se lze bezmyšlenkovitě podívat na strukturu anotací a jednoduchým porovnáním zjistit, v jakých částech docházelo k problémům a jaké části potřebují revizi. To je princip kvalitativního testování. Kvantitativní údaje, tj. metriky použitelnosti, však mohou poskytnout cenná doplňující data v přesných číslech. Čísla sice neřeknou, co přesně způsobuje problémovost této části produktu a jak tuto část vylepšit, ale pro prezentaci výsledků jsou solidní data přesvědčivější než výňatky slovních výpovědí s uživateli.

Skóre úspěšnosti (success score)

Metrika posuzující úspěšnost vykonaných úkolů bývá v řadě studií vždy na prvním místě. Pomocí této metriky lze zjistit, jak moc je daná část produktu chybová. Jestliže většina participantů odpověděla na otázku špatně, pak dále nemá smysl řešit detaily této části produktu a je nutné navrhnout úpravy pro zvýšení její použitelnosti. Úspěšnost (success) znamená, že se uživateli povedlo úspěšně splnit úkol, který mu byl zadán.

Základní vzorec pro výpočet skóre úspěšnosti:

$$\text{Skóre úspěšnosti} = \frac{\text{počet úspěšně dokončených otázek}}{\text{celkový počet všech otázek}}$$

Skóre úspěšnosti by mělo vycházet mezi 0 a 1 (nebo 0 a 100 %). Částečný úspěch je pak považován za neúspěch. Pro získání diferencovanějšího obrázku lze definovat i skupinu „částečně úspěšných“. Započítáním i signifikantnějších částečných úspěšných odpovědí lze získat celé spektrum. Je nutné se vyvarovat započítávání částečně úspěšných úkolů jako 0,5 hodnoty úspěchu a přičítání k celkovému průměru, neboť to zkresluje výsledky. Navíc lze říci, že skupina „částečně úspěšných“ úloh může prozradit více než celkové skóre úspěšnosti. Pomocí těchto skupin lze pochopit, v čem spočívá v dané části produktu problém. To je něco, co výzkumník od kvalitativního výzkumu očekává, zatímco kvantitativní výsledky by poskytly pouze úzce zaměřený soubor dat. Aby byl produkt považován jako produkt s dobrou použitelností, nemusí být skóre úspěšnosti 100 %. Průměrné skóre se pohybuje okolo 78 % (Paschenko, 2021).

Počet chyb (number of errors)

V uživatelské testování je za chybu (error) považována jakákoli chybná akce provedená při plnění úkolu. Podle Reasona (1990) existují dva typy chyb: tzv. „skluzu“ (slips) a omyly (mistakes). „Skluzu“ jsou provedené chyby se správným cílem (např. přechytnutí) a omyly jsou chyby s nesprávným cílem (např. špatné pochopení otázky -> špatná odpověď). Existují dva způsoby měření chyb: měření všech chyb (míra chybovosti) nebo zamření se na určitou chybu (míra výskytu chyb). Pro zjištění míry výskytu chyb je nejprve nutno vypočítat celkový počet chyb a vydělit je počtem pokusů. Je doporučováno počítat každou, i opakovanou, chybu. Chyby se lépe počítají v kombinaci s měřitelnými charakteristikami, např. počet kliknutí myši -> počet chybných kliknutí. U slovních výpovědí je zjištění počtu chyb náročnější a vyžaduje podrobnou analýzu a anotování zvukového záznamu. U zvukové nahrávky lze uvažovat jako chybu každé chybné slovo nebo označit celou větu jako chybnou, záleží na výzkumníkovi. Zvolenou metriku je ale nutno dodržovat u všech případů.

$$\text{Míra chybovosti} = \frac{\text{počet chyb}}{\text{počet všech pokusů}}$$

Míra chybovosti počítá se všemi typy chyb. Pro její výpočet je nutné předem definovat počty všech chyb, tj. všechny „skluzu“ a omyly. V závislosti na složitosti úlohy může být číslo větší nebo menší. Po definici všech chyb lze následně aplikovat vzorec pro výpočet míry výskytu chyb.

$$\text{Míra výskytu chyb} = \frac{\text{počet všech chyb}}{\text{počet všech možných chyb}}$$

Chybovost u všech produktů, i těch dokonalých, se téměř nikdy nerovná nule. Pravděpodobně neexistuje perfektní produkt, který by zabránil, aby lidé na něm netvořili omyly nebo jiné chyby. Z podstaty věci za to může lidská přirozenost. Pro člověka je přirozené tvořit chyby, proto je naprosto v pořádku, pokud se při testování použitelnosti chyby vyskytnou. Podle Saura (2010) pouhých 10 % úloh je dokončeno bezchybně a průměrný počet chyb na jednu úlohu odpovídá zhruba 0,7. Platí, že pokud se jedná o chybu, je třeba ji v další fázi eliminovat úpravou produktu.

Analýza chybovosti

Výzkumník si při vlastním anotování záznamu všímá, s jakými obtížemi splnili účastníci testu zadané úkoly. Každý úkol je následně výzkumníkem bodově ohodnocen podle náročnosti provedeného úkolu. Zvolena byla škála 0–2 body, kde

- 0 bodů znamená, že pro uživatele bylo snadné úkol provést,
- 1 bod znamená, že uživatel sice úkol splnil, ale s obtížemi nebo odpověděl jen částečně,
- 2 body znamená, že se uživateli nepodařilo úkol splnit.

Tabulku s ohodnocenou chybovosti je vhodné porovnávat s výsledky úspěšnosti vykonaných úkolů a skóre úspěšnosti. Pokud dochází u dané typy úlohy (dané části produktu) ke korelaci, pak lze označit tuto část jako potřebnou ke změně.

Task time (doba plnění úkolu)

Produkt s dobrou použitelností by měl být schopen poskytnout informaci uživateli rychle a úspěšně. Koncepce metriky pro měření doby plnění úkolu je jednoduchá, přesto musí být hodnotitelem využita efektivně. Nabízí se výpočet průměrné doby strávené respondenty nad jednotlivými úkoly.

Průměrná doba úkolu

$$= \frac{\text{doba prvního uživatele} + \text{doba druhého uživatele} + \dots + \text{doba } N \text{ uživatele}}{\text{počet všech respondentů}}$$

Výsledkem výpočtu je průměrná doba strávená dotazovanými nad úkolem. Je však výsledek dobrý nebo špatný? Pro určité jiné metriky existují jisté oborové standardy, ze kterých lze vycházet, ale pro čas řešení úkolu neexistují žádné. Přesto lze zjistit „ideální“ čas úlohy, a to podle výsledku zkušeného uživatele. U výběru reprezentativních uživatelů pro testování by měla být kompetentní osoba vzdělaná v daném oboru, ke kterému je

produkt určen, v tomto případě s kartografickým vzděláním. Čas strávený nad úkolem tohoto uživatele lze použít jako referenční. U hodnocení webového prostředí, např. webové mapy, lze sečíst průměrný čas pro každou akci, jako je „ukazování myši“ a „klikání myši“. Výpočet metriky doby plnění úkolu lze využít i jako archivační záznam pro porovnání s budoucí verzí produktu. Pokud průměrná doba u nové verze klesla, pak byl produkt s určitostí zlepšen. Ušetření času uživatele může mít velký dopad na celkovou kvalitu zážitku uživatele s produktem a také důvod, zda se uživatel k produktu vrátí nebo ze zkušenosti zvolí alternativní metodu zjištění požadované informace. Informace o době plnění úkolu lze také porovnat s úspěšností vykonaných úkolů a tím zjišťovat jejich signifikantnost.

Účinnost podle času (Time-based Efficiency)

Pro měření účinnosti existuje spousta způsobů. Jeden z nejzákladnějších je účinnost založená na čase, která kombinuje jak dobu plnění úkolu, tak skóre úspěšnosti, tzv. Time-based Efficiency.

$$\text{Time-based Efficiency} = \frac{\sum_{j=1}^R \cdot \sum_{i=1}^N \cdot \frac{n_{ij}}{t_{ij}}}{N \cdot R}$$

...kde R je počet uživatelů, N je počet úkolů, n_{ij} je skóre úspěšnosti úkolu i uživatelem j , t_{ij} je čas strávený uživatelem j nad úkolem i .

Poslední charakteristikou je zjištění spokojenost z hlediska uživatele s produktem. Toho lze docílit několika způsoby. Jedním z nich je vytvoření jednoduchého dotazníku po testování, a to buď formou tzv. SEQ (Single Ease Question) nebo SUS (System Usability Scale). Princip těchto dotazníků byl zmíněn v předchozích krocích.

Krok 10: Interpretace výsledků

Pomocí metrik použitelnosti lze vypočítat řadu hodnotných informací o testování a produktu. Vypočítané hodnoty je však nutné vhodně interpretovat a dávat do souvislostí. Také není od věci si pokládat dodatečné otázky, např. proč to tak vyšlo? Co to vlastně znamená? Je to vysoká nebo nízká hodnota? Odpovědi na takovéto typy dotazů je dobré vhodně uvést v závěrečné zprávě společně s navrženými možnými řešeními pro optimalizaci produktu.

Více podrobnějších informací včetně postupu ze dvou studií je k dispozici v textu diplomové práce.

Tomáš Vaníček