



Metodika měření multimediálního obsahu

MEDIARESEARCH, a. s.

15. 6. 2010

0. OBSAH

0. OBSAH	2
1. MODUL MĚŘENÍ MULTIMEDIÁLNÍHO OBSAHU	3
2. METODIKA	3
2.1. SBĚR DAT	3
2.2. MĚŘENÍ	4
2.2.1. KATEGORIZACE	4
2.2.2. REKLAMA	5
2.2.3. EMBEDOVANÝ OBSAH	6
2.3. VÝSLEDKY	6
2.3.1. ROZHRANÍ GEMIUSSTREAM	6
2.3.2. VEŘEJNÉ REPORTY	6
2.3.2.1. Návštěvy (Session)	7
2.3.2.2. RU	7
2.4. VERIFIKACE	9
3. BUDOUCNOST	10

1. MODUL MĚŘENÍ MULTIMEDIÁLNÍHO OBSAHU

Modul měření multimediálního obsahu (původně *streamovaného videa*) je součástí NetMonitoru, která je ve smlouvě se SPIR (Sdružení pro internetovou reklamu) definována následovně: „**Výsledky Výzkumu – streamované video** jsou výsledky získané realizátorem a obsahující informace o aktivitách uživatelů sledujících v rámci jednotlivých médií streamovaný obsah.“

Výsledky by měly být poskytovány následovně: „Realizátor je povinen předávat Výsledky Výzkumu – streamované video, tj. reporty z měření streamovaného videa Oprávněným osobám, a to ve formátu *.xls, popř. jiném vhodném formátu v rozsahu uvedeném v Metodice, vždy za dokončený kalendářní měsíc a to prostřednictvím oznamovacího emailu s informací o zpřístupnění dat na datovém úložišti (http resp. ftp server), vždy nejpozději do patnáctého kalendářního dne měsíce následujícího po kalendářním měsíci, na který připadá Produkční den pro ostatní typy dat za shodný měsíc měření, včetně; v případě, že dodání Výsledků připadne na Svátek, je dnem dodání nejbližší následující pracovní den.“

Vzhledem k tomu, že k výpočtu reálných uživatelů multimediálního obsahu jsou využívány rovněž výsledky z měření návštěvnosti serverů NetMonitor, bude mít produkce dat za multimediální obsah zpoždění oproti produkci dat NetMonitoru.

Cíle modulu společně s principy jsou definovány Streamingovou pracovní skupinou (SPS) podřízenou Metodologické komisí SPIR. Současným cílem je poskytovat komplexní informace o chování populace v oblasti sledování multimediálního obsahu (zvláště videa) uskutečňovaného v internetovém prohlížeči na měřených médiích. Jako vstupy používá výzkum data, která jsou sesbírána díky spolupráci měřených médií nebo třetích stran, které médiím dodávají systémy (např. web servery, streamovací servery). Výstupní data by měla poskytovat možnost srovnání s daty měření sledovanosti televize. Postupně by tedy mělo docházet ke sblížení nově ustanovené metodiky a metodiky TVmetrového měření. Společně s těmito fakty vidíme velkou důležitost v zapojení do výzkumu co nejvíce aktivních médií v České Republice.

Dodatečně SPIR rozhodl zapojit do výzkumu i měření Flashových online her. Zapojení každého takového média měřícího hry musí být zvlášť povoleno SPIR a realizátorem.

2. METODIKA

2.1. Sběr dat

Vstupní data jsou sbírána ve spolupráci s měřenými médii nebo jejich poskytovateli systémů (např. web serverů nebo streamovacích serverů). Způsob sběru dat závisí především na tom,

zda přehrávač umístěný na internetové stránce má otevřený zdrojový kód a zda ho médium nebo technický poskytovatel mohou modifikovat podle specifikace za účelem měření. Pokud je přehrávač založený na otevřených technologiích, implementuje se funkcionality reportování klientských akcí přímo do přehrávače. V případě jiných platform jsou data získávána z logů na serverech.

Technická implementace sběru dat je popsána v aktualizované technické specifikaci. V tuto chvíli garantujeme implementaci řešení pro Flash, MS Silverlight a Windows Media. O ostatních řešeních je v případě potřeby možná diskuze.

Aktuální dokumenty popisující technickou implementaci se dají najít na následující stránce:
files.stream.gemius.com

Login: gemiusStreaming

Heslo: gem12joml

2.2. Měření

Za účelem vypočítávání hodnot ukazatelů vyjmenovaných dále jsou shromažďována hrubá data, která obsahují serverový čas všech uživatelských akcí a meta data popisující obsah.

Metadata obsahují jednoznačnou identifikaci pořadu (ID pořadu) a délku daného pořadu (pokud ji pořad má). Další součástí metadat je také kategorizace (viz kapitola 2.2.1.).

Během přehrávání pořadu přehrávač předává měřicímu skriptu následující uživatelské akce:

- přehrávání (playing),
- pauza (paused),
- zastaveno (stopped),
- SeekingStarted (přetáčení),
- completed (konec materiálu),
- buffering (načítání obsahu).

Společně s těmito akcemi je odreportováno i v jakém úseku pořadu k této akci došlo. Na základě těchto údajů, se dá sestavit obrázek toho, jak uživatel daný materiál přehrával (jaké části přehrával, jaké přeskočil atd.). Jednotlivé uživatelské akce slouží k výpočtům hodnot tvrdých ukazatelů. Např. z délek přehraných úseků materiálu se vypočítává celkem strávený čas (TTS).

2.2.1. Kategorizace

Důležitou součástí meta dat je tzv. kategorizace materiálů.

Kategorizace se skládá z pěti povinných položek: NAZEV, TYP, ZDROJ, CIL, KATEGORIE. Volitelně může provozovatel využít kategorii META, kam si může nadefinovat libovolný uživatelský obsah.

Kategorizace je citlivá na velikost znaků. Prosíme tedy o dodržování konvence, kdy názvy položek jsou vždy velkými znaky, samotné hodnoty pak mohou mít smíšenou velikost znaků. U klíčových slov prosím dodržujte velikost písmen uvedenou v metodice.

NAZEV: Obsahuje jednoznačnou identifikaci videa; na základě názvu by mělo být snadné video najít pomocí lokálního vyhledávače.

TYP: Může nabývat hodnot „VOD“ pro video on demand; „linearni“ pro kontinuální vysílání (např. ČT24); „hra“ pro flashové hry; „reklama“ pro reklamní obsah.

ZDROJ: obsahuje název kanálu, ze kterého je pořad převzatý (ceskatelevize.cz může mít jako zdroj ČT24, ČT4 Sport atp.). Pokud je tato kategorie prázdná, bere se za zdroj médium, kterému patří měřicí skript.

CIL: obsahuje médium, na kterém obsah reálně běží (např. pro geewa.cz může být cílem hry.cz, centrum.cz atp.). Pokud je tato kategorie prázdná bere se za cíl médium, kterému patří měřicí skript. Tato položka také může obsahovat klíčové slovo „embed“, které značí, že obsah je embedován v neznámém nebo proměnlivém cíli.

KATEGORIE: Obsahuje třímístný kód, který v sobě nese informaci o obsahové kategorizaci materiálu a popisuje typ zdroje materiálu.

Pozice	Název	Popis	Příklady hodnot
1-2	Kategorie	Obsahová kategorizace doporučená kategorizační komisí SPIR	01 - filmy a seriály 02 - zpravodajství, publicistika 03 – dokumenty, vzdělávací pořady 04 - zábava, lifestyle, magazíny 05 – hudba, hudební pořady 06 - sport 07 - ostatní videa 08 - hry
3	Zdroj	Typ zdroje	1 - internetový obsah 2 - televizní obsah 3 - uživatelský obsah

Příklad: Seriál Ulice bude obsahovat v položce kategorie řetězec „012“. Pořad je tedy kategorizovaný jako filmy a seriály a jedná se o televizní obsah.

Technický popis, jak kategorizaci nasadit je k nalezení na stránce files.stream.gemius.com,

Login: gemiusStreaming

Heslo: gem12joml

2.2.2. Reklama

V první fázi není měřena videoreklama. Reklama vložená do pořadu je reportována jako akce 'paused' (při začátku reklamy) a 'played' (když se opět začne přehrávat pořad).

2.2.3. Embedovaný obsah

Embedované (vložené do cizího média) přehrávače je možné měřit, měřící skripty musí být vloženy společně s přehrávačem. Embedovaný přehrávač musí obsahovat gemiusStream Identifikátor (ID přidělené měřicímu účtu), cestu k hitkolektoru a všechny informace o materiálu (ID, kategorizace atd.). Tyto informace mohou být poslány jako parametry embedovaného objektu, nebo mohou být staženy přehrávačem při downloadování materiálu. Embedované přehrávače musí mít vypnutou funkci autoplay a návštěvnost videa se započítává médiu, které streamuje obsah embedovaného objektu (nikoli médiu, kde je vložený přehrávač). Embedovaný obsah musí být označen klíčovým slovem „embed“ v kategorii CIL. Výsledky měření embedovaného obsahu mohou být ve výstupech odděleny od výsledků naměřených na standardních (vlastních) přehrávačích.

2.3. Výsledky

Výsledky jsou prezentovány ve zvláštních reportech definovaných SPS, a dále pak v rozhraní gemiusStream, do kterého mají jednotlivá zapojená média přístup. V budoucnu by data mohla být prezentována i v aplikaci gemiusExplorer. Souhrny základních ukazatelů měření jsou produkovány a poskytovány SPIR na měsíční bázi souhrnně za celá média a po dohodě programové typologie také podle kategorií.

2.3.1. Rozhraní gemiusStream

Rozhraní gemiusStream (gS) je analogií rozhraní gemiusTraffic, které se používá v měření návštěvnosti NetMonitor: obsahuje online naměřená data, pro zapojené médium. Data jsou zde dostupná v měsíčním, týdenním, denním a hodinovém rozlišení. Statistiky za jednotlivá videa je možno získat na základně vyhledávání, skupiny videí lze analyzovat souhrnně na základě kategorizace. Reportované charakteristiky jsou podrobně popsány v dokumentu *Seznam_prezentovanych_ukazatelu_pro_multimedialni_obsah.pdf*.

Prezentace dat je v rozhraní gS opožděna cca 2 hodiny pro flashové (resp. obecně open code) přehrávače. Pokud jsou data dodávána prostřednictvím logů streamovacího serveru, data jsou zpravidla publikována následující den.

2.3.2. Veřejné reporty

Tato sekce obsahuje seznam ukazatelů, které budou reportovány ve veřejném reportu definovaném streamovací pracovní skupinou. Výsledky budou publikovány jedenkrát měsíčně ve formátu Excel souboru. Soubor bude obsahovat měsíční/týdenní/denní periody. Pro některé z ukazatelů budou k dispozici rovněž hodinová data (např. TTS, počet přehrání, ...). Datum pravidelné produkce tohoto reportu byl upřesněn v kapitole 1. Kromě uvedených ukazatelů jsou měřeny rovněž další charakteristiky (např. založené na cookies), které nebudou součástí veřejného výstupu, ale budou k dispozici zapojeným médiím v online rozhraní gemiusStream. Popis jednotlivých měřených charakteristik, včetně kompletního seznamu charakteristik prezentovaných v rozhraní gS je v dokumentu *Seznam_prezentovanych_ukazatelu_pro_multimedialni_obsah.pdf*.

Ukazatele, které jsou součástí veřejného reportu	
RU (počet reálných uživatelů)	Výpočet reálných uživatelů využívá obdobný algoritmus, jaký je používán v případě reálných uživatelů v NetMonitoru. Algoritmus je detailněji popsán v kapitole 2.3.2.3.
Návštěvy (Started Sessions)	Návštěva (Session) je série přehrání multimediálního materiálu generovaná uživatelem (Cookie). Mezi dvěma přehráními nesmí uplynout více jak 30 minut. Více v kapitole 2.3.2.1.
Zahájená přehrání (Views)	Zahájená přehrání ukazují, kolikrát bylo video nahráno do přehrávače a spuštěna první vteřina videa.
Unikátní přehrané materiály	Tento ukazatel udává počet unikátních přehraných materiálů ve sledovaném časovém období. To znamená, že, je-li materiál ve sledovaném období přehrán vícekrát, v tomto ukazateli je započítán jen jednou.
Total time spent (TTS) (celkový strávený čas)	Ukazatel uvádí, kolik času strávili uživatelé sledováním analyzovaného materiálu nebo skupiny materiálů.
TTS na RU (Time per RU)	Průměrná doba strávená reálným uživatelem sledováním multimediálního materiálu/skupiny materiálů. Počítá se jako: $\text{TTS na RU} = \text{TTS} / \text{RU}$ <p>Jedná se o speciální případ ukazatele ATS relative (využíván v TV metrovém výzkumu), kdy je uvažována cílová skupina všech uživatelů streamovaného obsahu.</p>

2.3.2.1. Návštěvy (Session)

Návštěva (Session) je série přehrání multimediálního materiálu generovaná uživatelem (Cookie). Mezi dvěma přehráními nesmí uplynout více jak 30 minut. Materiál je určený pomocí svého identifikátoru (pokud je v rámci 30ti minut stejný materiál přehrán na jiné stránce je započítán do stejné návštěvy; stále však musí jít o stejnou cookie).

Série přehrání stejného materiálu jednou cookie bude rozdělena do dvou Návštěv i v případě pokud:

- 1) Přehrávání je ukončeno pauzou a v následujících 30ti minutách se neobjeví událost playing.
- 2) Během přehrávání se objeví bufferování, které trvá více jak 30 minut.

2.3.2.2. RU

V rámci měřených dat máme k dispozici kompletní informace pro jednotlivé cookies. Nicméně tento ukazatel dosahuje výrazně větších hodnot, než skutečný počet reálných ukazatelů, neboť odhad je zatížen problematikou mazání cookies, používáním více počítačů a prohlížečů, atd. Obdobně jako v případě měření návštěvnosti serverů je nutné využít algoritmus, který převede údaje o cookies na údaje o reálných uživatelích streamovaného

obsahu. V případě streamingu je mezi reálné uživatele zahrnut každý, kdo odsledoval jakoukoli část streamovaného obsahu v daném časovém intervalu. Pokud bychom tedy porovnávali reach uživatelů streamovaného videa, tak bychom v rámci TVmetrového měření měli uvažovat sekundový reach.

Odhad **reálných uživatelů** je založen na estimated-cookie algoritmu (ve výpočtu bude využívána koncepce good cookies s 15-ti denním čekáním). Jedná se o analogii k výpočtu reálných uživatelů při měření návštěvnosti stránek NetMonitor. Ke konverzi Cookies na reálné uživatele je využíván ukazatel přehrání s přesahy – Stream (podobně jako jsou PV používány pro výpočet EC v NetMonitoru).

Přehrání (Stream) je nepřerušené přehrání kompaktní části materiálu. Přehrání začíná událostí Playing a bývá ukončeno některou z následujících událostí: SeekingStarted, paused, buffering, stopped, completed. Protože počet přehrání odkazuje na počet uživatelských akcí, které započnou přehrávání materiálu, rozhodla SPS některé ukazatele odvozené od přehrání nazvat *počet uživatelských akcí*. Jedná se o jednu ze základních jednotek měření. V případě výpočtu reálných uživatelů je využíván ukazatel Počet přehrání s přesahy, který uvažuje i přehrání, která byla započata před monitorovaným obdobím, ale jejich trvání přesáhlo i do monitorovaného období (v této kapitole budeme značit jako přehrání).

Algoritmus využívá informace o chování cookies, které jejich „vlastníci“ (uživatel počítače) nemažou, o chování cookies, které jsou mazány, a dále informace z návštěvnosti internetu měřené v projektu NetMonitor. Využití těchto údajů umožňuje poměrně přesně určit počty reálných osob – návštěvníků multimediálního obsahu.

Algoritmus pro reportování reálného uživatele pracuje ve dvou krocích:

- 1) Výpočet počtu cookies (tzv. estimated cookies), které by systém naměřil na daném médiu S zapojeném do měření multimediálního obsahu, kdyby neexistoval problém mazání cookies.

Systém analyzuje vztah počtu všech cookies a počtu všech přehrání vůči počtu tzv. validních cookies (označované rovněž jako good cookies) a počtu jimi generovaných přehrání (tzv. Good streams). Za validní cookies jsou považovány takové cookies, u nichž systém zaznamenal aktivitu před měřeným měsícem, v něm a rovněž do 15 dnů po jeho skončení. Z výše uvedených charakteristik lze vypočítat počet cookies, které by systém naměřil, kdyby bylo možné pomocí cookies dlouhodobě a jednoznačně označit všechny počítače připojující se k internetu.

$$EC_S = \min [All\ cookies_S; Good\ cookies_S * (All\ streams_S / Good\ streams_S)], \text{ kde:}$$

EC – Estimated Cookies na médiu/serveru/sekci *S*

All cookies – všechny naměřené cookies

Good cookies – počet nesmazaných (validních) cookies

Good streams – přehrání generovaná nesmazanými (validními) cookies

All streams – všechna přehrání generovaná všemi cookies

Funkce minima se používá pro zamezení situace, kdy by počet EC překročil všechny naměřené cookies.

2) Stanovení počtu reálných uživatelů pro dané médium/server/sekci *S*:

V předchozím bodě získáváme hodnotu *EC* naměřenou pomocí metodiky multimediálního obsahu. Obdobně – za použití pageviews – je v rámci návštěvnosti internetu počítán odhad *EC* v NetMonitoru (označme hodnotu měsíční hodnotu *EC* za všechny měřené servery NetMonitor jako $EC_{NetMonitor}$). Počet reálných uživatelů multimediálního obsahu získáme z následujícího vzorce:

$$RU_S = EC_S * coeff_{NetMonitor}, \text{ kde:}$$

RU_S – počet Reálných uživatelů pro dané médium/server/sekci

EC_S – Estimated Cookies na médiu/serveru/sekci *S* počítaný podle bodu 1)

$coeff_{NetMonitor}$ – koeficient z měření NetMonitor, který se počítá jako:

$$coeff_{NetMonitor} = Univerzum_{NetMonitor} * EC_{NetMonitor} .$$

Uvedený postup je analogický s postupem, který je využíván v případě výpočtu reálných uživatelů v NetMonitoru.

Výpočet denních a týdenních *RU* probíhá dle obdobného mechanismu jako výpočet měsíčních hodnot. Denní a týdenní *EC* pro jednotlivé servery se počítají podle výše uvedeného vzorce.

2.4. Verifikace

Verifikace je zásadní moment implementace sběru dat. Hlavní důvody verifikace jsou dva:

- ❖ Technologie je sama o sobě nová a stále je třeba ji testovat, zvláště při nasazení v nových podmínkách.

- ❖ Implementace funkcionality pro reportování akcí na hitcollector je mnohem komplexnější, než je to v případě implementace standardních skriptů NetMonitoru.

Každý provozovatel, který je zapojený do výzkumu by měl projít fází verifikace předtím, než jeho data budou oficiálně měřena.

Vzhledem k výše uvedenému vznikla pravidla pro testování, která potvrdí správné nasazení měření na serveru. V tuto chvíli mají pravidla tuto formu:

- ❖ Každý spolupracující subjekt je nutné testovat zvlášť.
- ❖ Verifikace je prováděna na dvou úrovních – kvalitativní a kvantitativní.
- ❖ Kvalitativní verifikace (popsaná detailně v separátním dokumentu – NM055) znamená simulovat předem definované uživatelské akce na straně realizátora na testovacím přehrávači/stránce/materiálu (před tím, než budou změny implementovány do produkční verze klienta/streamovacího serveru) a sledování správnosti sbíraných dat přímo z logů na hitcollectoru, nebo při odesílání v rámci prohlížeče (klientského počítače).
- ❖ Kvantitativní verifikace spočívá v získání interních statistik média za dostatečné časové období (více jak týden) a za skupinu videí (v produkčním režimu masivní přístupy) a porovnat je s nasbíranými daty.
- ❖ Zahájená přehrávání (načtení videa do přehrávače a uživatelská akce *play*) nebo TTS jsou dobří kandidáti pro účely kvantitativní verifikace. Metodika výpočtu srovnávaných ukazatelů by měla být shodná.
- ❖ Pokud se výsledky měření nebudou lišit o více než 10% v porovnání s interními ukazateli média, budeme médium považovat za ověřené
- ❖ Kvantitativní testy budou vykonány pro servery Stream, iDNES, Novinky.cz a Česká televize. V dalších případech nově zapojených médií budou na straně realizátora prováděny pouze kvalitativní testy.

3. BUDOUCNOST

Níže je výčet předpokládaných změn a vylepšení v projektu:

- Měření video reklam zobrazených v rámci přehrávání videa
- Publikování ukazatelů GRP a rating pro vybrané pořady nebo skupiny pořadů
- Realizace obsahové kategorizace multimediálního obsahu na straně médií
- Provázání návštěvnosti se sociodemografií uživatelů