

Analytika mluveného slova

výňatek z technické specifikace

Obsah

Analytika mluveného slova.....	1
1. Uvedení do situace	2
2. Základní informace a stručný popis zakázky	2
2.1 K čemu transkripci a identifikaci osob ČRo využije?.....	3
2.2 Předpokládaný vývoj systému	4
2.3 Co je úkolem Dodavatele, hlavní body	6
3. Živé vysílání – Dante protokol, zdroj pro real-time AMS.....	6
3.1 Technologie:	6
3.2 Místo k připojení:	6
4. Serverové řešení.....	6
4.1 Základní parametry serverů:	6

Analytika mluveného slova

1. Uvedení do situace

Český rozhlas (ČRo) hodlá zpřístupnit některé obsahy vysílání v textové podobě pro snadné hledání i sledování obsahu živého vysílání v textové podobě. Pro tento účel je nutné vyvinout Analytiku mluveného slova (AMS) rozepsanou níže tak, aby bylo možné ve stejné kvalitě zpracovat jak živé vysílání, tak archiv kontinuálního vysílání kterým ČRO disponuje.

ČRo hodlá AMS zavést nejprve pro živé vysílání a to v reálném čase dle technicko-technologických možností s latencí transkripce v řádu sekund či jejich nižších desítek v případě identifikace. Takto zpracovávané vysílání bude automaticky ukládáno, kategorizováno a doplněno metadaty a může být po skončení každého z pořadu znovu důkladněji analyzováno.

ČRo dále disponuje archivem kontinuálního vysílání i jeho základním popisem metadaty u některých stanic už od roku 2003. Záměrem je archiv kontinuálního vysílání analyzovat a převést veškeré mluvené slovo na text (transkripce), rozlišit mluvčí (diarizace), identifikovat mluvčí, promluvy přiřadit k jejich profilům a indexovat je pro snadné vyhledávání. Tento archiv bude navazovat na zpracovávané živé vysílání.

Třetím předmětem zakázky je dodání systému pro on-demand AMS v plné i zcela jednoduché podobě.

AMS je druhou polovinou celkové analýzy kontinuálního vysílání, jejíž první částí je Analytika vysílání (AV), ve které se zvuková stopa člení na jednotlivé úseky a tyto označuje. Celkovým výsledkem, ve kterém dochází k propojení obou částí, tedy AMS a AV, je podrobný popis kontinuálního vysílání, tedy včetně transkripce s případnou identifikací mluvčích ve vysílání a to všech stanic ČRo. A tento výstup, tedy jak zpracovaný archiv tak online vytvářený archiv přímo ze živého vysílání, bude zpřístupněn dle práv uživatelům na internetu v rámci projektu mujROZHLAS.

mujROZHLAS je strategickým projektem ČRo a bude největším českým audio archivem na internetu. Předpokládané období pro spuštění první verze projektu je plánováno na jaro 2019.

Český rozhlas nehledá pouze dodavatele, ale aktivního a spolehlivého partnera pro vývoj jedinečného produktu, kterému půjde o dosažení co nejlepšího možného výsledku a nezalekne se nesnadného vývoje.

2. Základní informace a stručný popis zakázky

Zakázka obsahuje:

1. Vývoj, dodání, instalaci a aktualizaci systému AMS, čímž je míněna transkripce, diarizace a identifikace, výroba speciálních slovníků a voiceprintů rozepsaná podrobněji níže, pro analýzu
 - živého vysílání,
 - on-demand audiosouborů,
 - archivu kontinuálního vysílání.
2. Hromadné zpracování archivu kontinuálního vysílání systémem AMS na straně dodavatele (ca 100 TB audiosouborů v MP3).

3. Servis systému po dobu trvání zakázky.

Základní informace k AMS živého vysílání, on-demand a archivu

- AMS automaticky rozpoznává mluvené slovo (ASR), identifikuje jazyk (čeština, slovenština, ostatní – tedy dva a ostatní zatím nerozlišuje), provádí transkripci a přiřazuje timecode tak, aby bylo možné přistupovat ke konkrétnímu úseku zvukové stopy dle času.
- Systém dále rozlišuje mezi mluvčími a pohlavím (muž/žena) v případě neidentifikování mluvčího, a pokud je to možné, identifikuje mluvčí tak, aby ČRo mohla přiřadit promluvu k hlasovému profilu osoby. Profil by měla každá osoba promlouvající ve vysílání, i když může jít o anonymní profil, který lze posléze identifikovat.
- Pro vybrané osoby (v řádu jednotek, např. Miloš Zeman, Andrej Babiš) připraví dodavatel personalizované slovníky takové, aby přesnost transkripce byla v rozmezí 95-100 %. V rámci technických možností připraví dodavatel ve spolupráci s ČRo specializované – tematické – slovníky, např. pro vědecké pořady ČRo, případně modul pro sníženou kvalitu hlasu (telefonát, rušné místo, sportovní přenos).
- Tento systém může používat pro identifikaci mluvčích jak obvyklá schémata pořadů a jejich tvůrců dostupná z metadat popsaných níže, tak samotnou transkripci, ve kterém se obvykle identifikuje následující mluvčí v promluvě. Základem pro identifikaci bude digitální otisk hlasu, tzv. voiceprint, které bude dodavatel rovněž vytvářet.
- Provozování systému pro živé vysílání i pro on-demand požadavky bude na serverech ČRo.
- Hromadné zpracování archivu kontinuálního vysílání provede dodavatel na svých (pronajatých) zařízeních. Data pro AMS archivu dodavatel převezme osobně na svých zařízeních (ca 100 TB dat).
- Pro jednotlivé nebo menší celky připraví dodavatel pro ČRo on-demand systém, který umožní opětovné zavedení výsledků analýzy do databází ČRo. On-demand AMS má funkce volitelné, tedy je možné požadovat např. pouze transkripci.
- Nákladnost na nákup HW, příslušného OS a implementaci je součástí hodnocení.
- Podrobný popis níže popisuje i části, které dodavatel nevyvíjí, ale např. pouze poskytuje data pro jejich plnění.

2.1 K čemu transkripci a identifikaci osob ČRo využije?

2.1.1 Indexace obsahu

Indexace a vyhledávání v audio – pro interní a kategorizační / archivní účely i pro vyhledávače na internetu a jeho uživatele. Indexaci a vyhledávání řeší ČRo jinou veřejnou zakázkou.

2.1.2 Sledování vysílání i bez zvuku

Posluchači umožní sledovat přepis online v případě, když nemůže poslouchat. Umožní tak nepřetržité sledování informací v rámci možností kvality AMS. Online sledováním se míní přepis mluveného slova se zpožděním v sekundách od odvysílání s diarizací a identifikací mluvčího atd.

2.1.3 Indexace obsahu vyhledávači.

Současné vyhledávače neumí prohledávat zvukovou stopu, transkripce jim umožní přečíst si obsah audia a nechat jej zaindexovat. ČRo tak zpřístupní vybrané obsahy svého vysílání širší veřejnosti i prostřednictvím vyhledávačů.

2.1.4 Tematizace obsahu

Dalším krokem je automatická tematizace / štítkování obsahu, které umožní rychlejší kategorizaci.

2.1.5 Automatická shrnutí

Transkripce bude časem sloužit pro vytváření automatických shrnutí obsahu - sumarizace.

2.1.6 Speech to text

Jakmile bude možné vytvářet sumarizace, ČRo bude pracovat na automatickém převodu textu na hlas.

2.2 Předpokládaný vývoj systému

Na transkripci češtiny a slovenštiny se pracuje již desítky let, a proto ČRo nepředpokládá, že by použité technologie pro transkripci měly být zcela nově vyvinuty, nicméně ani takový přístup nevyklučuje. Níže jsou uvedené use case, které by měl systém plnit v nejlepší možné kvalitě. Úkolem dodavatele je nejen vyvíjet všechny uvedené use case, ale zároveň pracovat na jejich vyšší přesnosti a spolehlivosti.

2.2.1 Use case

Dodavatel může vyvinout systém fungující na jiných krocích s přesnější analytikou než je níže uvedený výčet, nicméně ČRo tyto use case považuje za základní hodnotící kritéria.

ČRo očekává (ale nevyžaduje, pokud dodavatel nabídne lepší řešení) vývoj systému AMS rozdělený na tyto navazující úseky:

1. Příjem vysílání (signálu) ano / ne
 2. Ticho / zvuk
 3. Identifikace promluvy ve zvuku
 1. Promluva
 2. Identifikace jazyka promluvy (defaultně počítá s češtinou, a slovenštinou jako druhým jazykem) / jiné zvuky
 4. Promluva – její transkripce / ostatní zvuky v promluvě¹
 1. Správný počet slov včetně předložek
 2. Počet správně rozpoznaných slov
 3. Správný počet přesně přepsaných slov
 4. Přesnost na timecode
 5. Odfiltrování obvyklých hlasových i mimoslovních projevů, např. přitakání, zakoktání, přeřeknutí a opakování
 6. Správný počet přesně určené interpunkce
 5. Transkripce telefonního vstupu a vstupů se sníženou kvalitou zvuku (např. diktafon)
 6. Diarizace
 7. Identifikace pohlaví (muž / žena) v případě neidentifikování jedince
 8. Identifikace jedince
 9. Crossover² promluvy (dva a více hlasů přes sebe, hlas s ruchem promluv v pozadí)
 10. Transkripce s ruchem na pozadí (sportovní přenos)
 11. Vytváření úseků⁴
 12. Přiřazení práv úsekům⁴
 13. Sloučení úseků do vyšších celků⁵
 14. Přiřazení práv vyšších celků⁴
-

Poznámky:

¹ Transkripce promluvy – lidského hlasu, kvalita transkripce se hodnotí dle kritérií viz níže.

² Crossover je situace běžná ve vysílání, ve které přes sebe jdou dva hlasy, hudba či zvuková grafika. Údaje o začátcích a koncích úseků mluvčích se tedy budou překrývat.

³ Rozlišení nového úseku bez oddělovací znělky / jinglu. Např. Zprávy v celou hodinu se skládají z promluvy zprávaře a příspěvků redaktorů. Bude řešeno i s pomocí Analýzy mluveného slova.

⁴ Provede ČRo.

⁵ Provede dodavatel a ČRo, přesné vymezení vyplyne z vývoje systému.

2.2.2 Přesnost analytiky

ČRo považuje za rozhodující v přesnosti transkripce a analytiky níže uvedená kritéria AMS, jejich váha pro hodnocení je součástí zadávací dokumentace. Kritéria pro hodnocení jsou součástí zadávací dokumentace veřejné zakázky:

1. Rozpoznání mluveného slova
2. Rozpoznání jazyka (čeština, slovenština, ostatní)
3. Správný počet slov včetně předložek
4. Počet správně rozpoznaných slov
5. Správný počet přesně přepsaných slov
6. Přesnost na timecode
7. Odfiltrování obvyklých hlasových i mimoslovních projevů, např. přitakání, zakoktání, přeřeknutí a opakování
8. Správný počet přesně určené interpunkce
9. Správný počet přesných určení mluvčího – diarizace (kolikrát která osoba mluvila)
10. Rozlišení mluvčích muž/žena (ano/ne)
11. Správný počet identifikovaných mužů / žen
12. Správný počet identifikovaných mluvčích (kolik unikátních osob mluvilo)
13. Crossover transkripce (dva a více mluvčích hovoří najednou)

2.2.3 Přesnost a lidský faktor

Ideální systém by měl analyzovat vysílání bezpečně bez zásahu člověka. Ač je cílem 95% a vyšší přesnost transkripce, ČRo nepředpokládá takovou funkčnost v nejbližším období. ČRo předpokládá, že bude nutné analytiku ladit na straně dodavatele. V případě, že systém dodavatele vyžaduje stálou lidskou pracovní sílu na straně ČRo, musí ji specifikovat a uvést do ceny AMS.

2.2.4 Analýza vysílání - propojení

Pomocí Analýzy vysílání (AV), která je předmětem jiné veřejné zakázky, bude ČRo kategorizovat obsah vysílání, určovat začátky a konce pořadů, znělky, promluvu apod. ČRo předpokládá, že bude možné na své straně zpřesňovat kategorizaci, např. jako kontrolní zdroj pokud se nepodaří AV rozlišit reportáž s promluvou redaktora a promluvu moderátora.

2.2.5 Harmonogram a vývoj

Závazný harmonogram je uveden v příloze.

Stručně z harmonogramu: V první etapě předpokládáme nadefinování HW a zavedení online AMS v maximálně možné míře dle 2.2.1 a to do konce března 2019. Rozpracování use casů a jejich

postupné nasazování do provozu či zpřesňování, nasazení AMS pro on-demand zpracování audií, slovníky atd. i pro zpětné zpracování audií z vysílání již prošlých AMS (opětovné zpracování archivu).

2.2.6 Systém

Dodaný systém AMS bude ve vlastnictví ČRo.

2.3 Co je úkolem Dodavatele, hlavní body

- Dodavatel dodává, instaluje, aktualizuje, vyvíjí a zkvalitňuje software pro AMS v úzké spolupráci s ČRo dle uvedených Use case (2.2.1) jak pro živé vysílání tak pro on-demand.
- Dodává a upravuje hlavní slovník, slovníky tematické i personalizované.
- Dodává data pro hlasové profily (popis níže).
- Dodavatel dodá a vyvíjí API dle níže uvedených požadavků v úzké spolupráci s ČRo.
- Dodavatel nadefinuje HW a OS, který ČRo zakoupí a bude na něm systém provozovat.
- Dodavatel průběžně implementuje a aktualizuje software AMS systému na HW v ČRo.
- Dodavatel poskytuje po dobu určenou support systému, v začátcích a při nasazování nových modulů 24/7 a posléze 12/7.
- Dodavatel bude vytvářet hlasové otisky – voiceprint pro uložení / zavedení do hlasového profilu osoby, který bude spravovat ČRo a které bude využívat pro identifikaci mluvčích.

3. Živé vysílání – Dante protokol, zdroj pro real-time AMS

Pro AMS živého vysílání se musí dodavatel podřídit technickým podmínkám ČRo. Připojení na vysílání dodavatelem je možné pouze skrze rozhraní protoklu Dante v serverovně ČRo v a provádění analytiky živého vysílání je tek možné pouze sídle ČRo:

3.1 Technologie:

- Standard: AoIP – audio over IP
- Protokol: Dante (<https://www.audinate.com/products/software/dante-virtual-soundcard>)
- Vzorkovací frekvence: 16-bit audio ve frekvenci 48 kHz
- Vysílání: modulace pro jeden program je stereo
- V systému Dante je celkem 64 kanálů, tj. až 32 stereo programů

3.2 Místo k připojení:

- Serverovna v sídle Českého rozhlasu, Vinohradská 12, 120 99 Praha 2, ČR
- Serverové řešení, základní parametry viz bod 7.
- Pro testování jsou dostupné použít streamy ČRo

4. Serverové řešení

Systém bude provozován na serverech v budově Českého rozhlasu, Vinohradská 12, Praha a by měl splňovat parametry níže uvedené. Ostatní nadefinuje dodavatel řešení.

4. 1 Základní parametry serverů:

- Izolované servery
- Rackové provedení, 2 zdroje
- Preferujeme Intel platformu standardních výrobců HP, Dell, Lenovo...
- 2x dedikované ethernetové porty DANTE HW kartu

- 2x ethernetové porty Dante pro síťové propojení
- Operační systémy: Windows, Linux
- Počet procesorů: 2/4
- Připojení SAN 8/16 Gb, zdvojené
- Síťovém připojení 1/10 Gb, zdvojené
- Napájení: 230 V
- Vzdálená správa
- Zdvojené řešení s minimální dobou výpadku:
- Řešení, které umožňuje zpracovat vysílání z archivu – v případě výpadku serverů
 - Minimální doba výpadku je
 - Otevřené řešení pro možnost doplnění clusterovým řešením