eScribe: ONLINE PŘEPISOVACÍ CENTRUM

Zdeněk BUMBÁLEK, Jan ZELENKA, Ivan KUTIL
Research and Development Centre, Fakulta elektrotechnická
České vysoké učení technické v Praze
Technická 2, 166 27, Praha 6
Email: {bumbazde, zelenj2}@fel.cvut.cz
xkuti00@vse.cz

Anotace:
Komunikační a hlasové technologie mají potenciál snížit komunikační, vzdělávací a sociální bariéry osob s postižením sluchu. Cílem projektu eScribe je omezení těchto bariér a pomocí zapojení moderních komunikačních technologií (mobilní a sociální sítě, Internet, internetová telefonie, automatické rozpoznávání řeči) přispět ke všeobecné dostupnosti přepisu mluvené řeči.

1. Úvod

Přes velké pokroky v rozvoji asistivních a zejména hlasových služeb (rozpoznávání a syntéza řeči) v posledních letech je jejich dostupnost pro komunitu hendikepovaných stále značně omezená. Na jedné straně existují bariéry související s komplikovaností a nákladností těchto nových technologií, na straně druhé je jejich dostupnost omezená kvůli centralizovaným a lokálním řešením. Přesto, případné široké zpřístupnění těchto technologií by mělo dramaticky pozitivní dopad na kvalitu života hendikepovaných.

Projekt eScribe [13] je společnou aktivitou výzkumného vývojového centra RDC ČVUT [14] a České unie neslyšících [15] a zabývá se zpřístupněním asistivních hlasových služeb (rozpoznávání a okamžitá transkripce řeči do textu, zprostředkování textové a grafické informace hlasem aj.) široké komunitě uživatelů z řad hendikepovaných pomocí moderních komunikačních technologií. Cílem je navrhnout architekturu a vytvořit prototyp univerzální webové služby, použitelné přes různorodé běžné mobilní terminály, zpřístupňující tyto asistivní hlasové služby hendikepovaným kdykoli a kdekoli. Projekt se ve své první fázi zaměřil na standardní řešení na základě architektury klient-server a převodu řeči do textu za pomoci speciálně vyškolených přepisovatelů [1]. Současná druhá verze řešení je založena na principu nastupujícího paradigmatu Cloud Computing, který značným způsobem snižuje nároky na vlastní HW i SW poskytovatele služby a umožňuje mu tak poskytovat službu s nižšími náklady a tím i většímu počtu potencionálních uživatelů. Ve druhé fázi projektu došlo také k propojení komunikační architektury s automatickým systémem rozpoznávání řeči společnosti NEWTON Technologies, a.s. – partnera projektu [16]. Za účelem zvýšení úspěšnosti rozpoznávání bude projekt eScribe využívat služeb tzv. stínového mluvčího, který pracuje jako prostředník mezi uživatelem a systémem pro rozpoznávání hlasu. Práce stínového mluvčího je mentálně velmi náročná činnost vyžadující značné zkušenosti a dlouhodobý trénink. Díky těmto osobám je však možné dosahovat podstatného zvýšení úspěšnosti rozpoznávání bez nutnosti dalších korektur [5]. Obdobně jako byly limitujícím faktorem první verze projektu náklady a dostatečný počet vyškolených přepisovatelů, je druhá verze projektu omezena počtem a náklady na stínové mluvčí. Možným řešením omezení obou verzí je crowdsourcing přepisovatelů a tlumočníků v rámci sociálních sítí. Tento nový přístup řešení asistivních služeb je představen v závěrečné kapitole věnované budoucí práci projektu.

2. Vědecká závažnost a aktuálnost

V České republice žije podle odhadů ČUN cca 300–500 tis. občanů s postižením sluchu, dalších 100 tis. pak tvoří lidé slabozrací nebo zcela nevidomí. Komunikační a hlasové technologie mají potenciál snížit komunikační, vzdělávací a sociální bariery handicapovaných spoluobčanů. S nastoupeným trendem stárnoucí populace ČR a EU budou asistivní technologie nabývat stále vyššího významu. Asistivní hlasové služby založené na ASR (Automated Speech Recognition) a TTS (Text-to-Speech) jsou předmětem výzkumu předních světových univerzitních pracovišť, což dokladují četné vědecké publikace. Architektury těchto služeb jsou však často centralizované a založené na proprietárním řešení, což limituje vzájemnou propojitelnost těchto systémů a

jejich dostupnost prostřednictvím standardních koncových zařízení. V důsledku toho jsou pak nabízené služby využívány pouze zlomkem uživatelů z potenciální cílové skupiny. Cílem projektu je toto omezení odstranit a pomocí zapojení moderních komunikačních technologií (mobilní sítě, Internet a internetová telefonie) přispět ke všeobecné dostupnosti těchto platforem.

Z vědeckého hlediska lze k modelování architektur asistivních hlasových služeb přistupovat dvěma způsoby. První přístup (eScribe verze 1) je založen na dobře známém modelu webových služeb klient-server, jehož využívání pro široce dostupné asistivní služby je, i přes zřejmý potenciál, dnes stále velice omezené. Novátorský přístup pak představuje model označovaný v anglické literatuře jako „Cloud Computing (CC)" konkrétně „Platform as a Service (PAAS)". CC představuje moderní přístup k modelování technologií a služeb založených na Internetu a je předmětem řady vědeckých publikací. Organizovaně se danou problematikou zabývá fórum CCIF (Cloud Computing Interoperability Forum) sdružující přední světové IT organizace. Využití CC v asistivních aplikacích představuje nový a inovativní přístup k dané problematice. Zcela novým řešením problematiky je záměr využití potenciálu sociálních sítí a crowdsourcing asistivních služeb v rámci těchto sítí.

3. Současný stav řešeného problému

Asistivní hlasové služby jsou založeny na ASR a TTS technologiích a popsány jsou v řadě studií [5][6]. Pro rozpoznávání anglického jazyka existuje v současnosti již i řada komerčních řešení (např. Microsoft, IBM). Vedoucí postavení ve využití ASR v asistivních technologiích zaujímá LLP (Liberated Learning Project [9]) ve spolupráci s IBM Human Ability and Accessibility Center [11]. Rozpoznávání národních jazyků jako je čeština zůstává mimo zájem velkých společností a rozvíjí se především na univerzitní půdě (v ČR: TU Liberec – prof. Nouza a kol., ZCU – doc. Müller a kol., ČVUT – Speech Processing Group). Na ČVUT se zapojením ASR a TTS v asistivních technologiích zabývají projekty RDC eScribe [1] a Voice2Web [3]. Problematika Cloud Computing (sdílení hardwarových i softwarových prostředků pomocí sítě) je předmětem výzkumu předních světových IT společností a univerzitních pracovišť (Kadar M., 2009), (Linthicum, D., S., 2009). Využití CC v asistivních aplikacích však představuje nový a inovativní přístup k dané problematice.

4. Obecný princip přepisu poskytovaného na dálku

Technické řešení projektu eScribe je založeno na IP telefonii a online zobrazení přepisu řeči na webových stránkách či mobilní aplikaci. Přístup do systému je možný jednak z klasických telefonních přístrojů, mobilních telefonů, ale i z HW SIP telefonů a SW SIP klientů. Nejsnadnější způsob pak tvoří webový klient, díky kterému je celý systém dostupný pouze z webového rozhraní bez nutnosti jakékoli instalace a konfigurace pro uživatele.

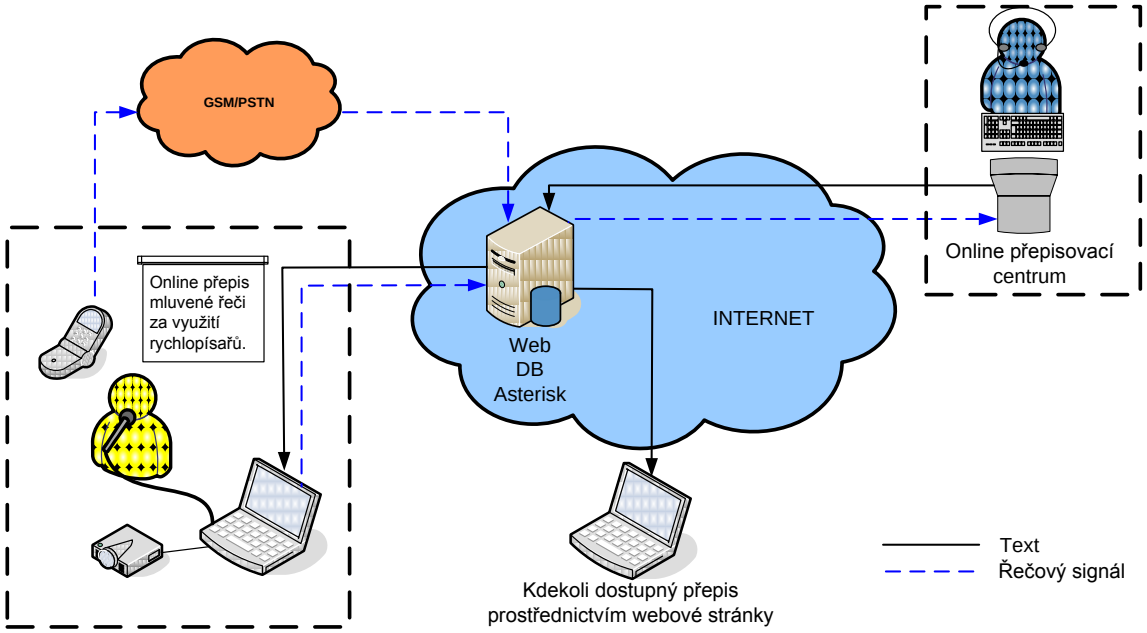
Samotný proces přepisu je možné realizovat 5 způsoby:

1. Prostřednictvím služeb speciálně vyškolených přepisovatelů fyzicky přítomných na místě konaného přepisu [15]
2. Prostřednictvím online dostupných služeb speciálně vyškolených přepisovatelů [1]
3. Využitím systémů automatického rozpoznávání řeči (ASR) [7][8]
4. Kombinace bodů 2 a 3 (přepisovatel pracující jako korektor) [10]
5. Doplnění systému ASR o služby stínového mluvčího [5]

5. eScribe I – prototyp řešení klient-server

5.1 Princip a funkce

Princip původní vývojové verze systému eScribe I je znázorněn na obr. 1. Z místa konání přednášky pro neslyšící je přenášen hovor pomocí VoIP telefonie do přepisovacího centra nebo na jiné místo, kde se nachází přepisovatel. Přepis zajišťují speciálně vyškolení rychlopisaři, kteří používají velký seznam zkratek. Ten se expanduje na celá slova nebo věty. K tomu využívají MS Word a funkci automatického vkládání pro vložení těchto zkrácenin do dokumentu. Použití textového editoru bylo jednou ze základních podmínek přepisovatelů na vývoj aplikace pro přepis. V praxi probíhá přepis tak, že si přepisovatel pomocí webového rozhraní aplikace vytvoří a otevře dokument MS Word. Na jeho pozadí běží programový kód, který s malým zpožděním, prakticky v reálném čase, odesílá text na server, odkud je dále zobrazován na webovou stránku. Na místě přednášky je k dispozici projektor a přepisovaný hlas je zobrazován na plátno. [2].



Obr.1: Princip systému eScribe I

6. eScribe II – současné řešení online přepisu v prostředí „cloud computing“

6.1 SW ústředna Asterisk

Asterisk [4] je pobočková softwarová ústředna s otevřeným zdrojovým kódem a je určena k provozování telefonních služeb jak na úrovni přepínání okruhů (TDM) tak v síti s přepínáním paketů. Systém Asterisk lze právem považovat za velmi rozšířené, flexibilní a silné řešení v oblasti telekomunikačního SW.

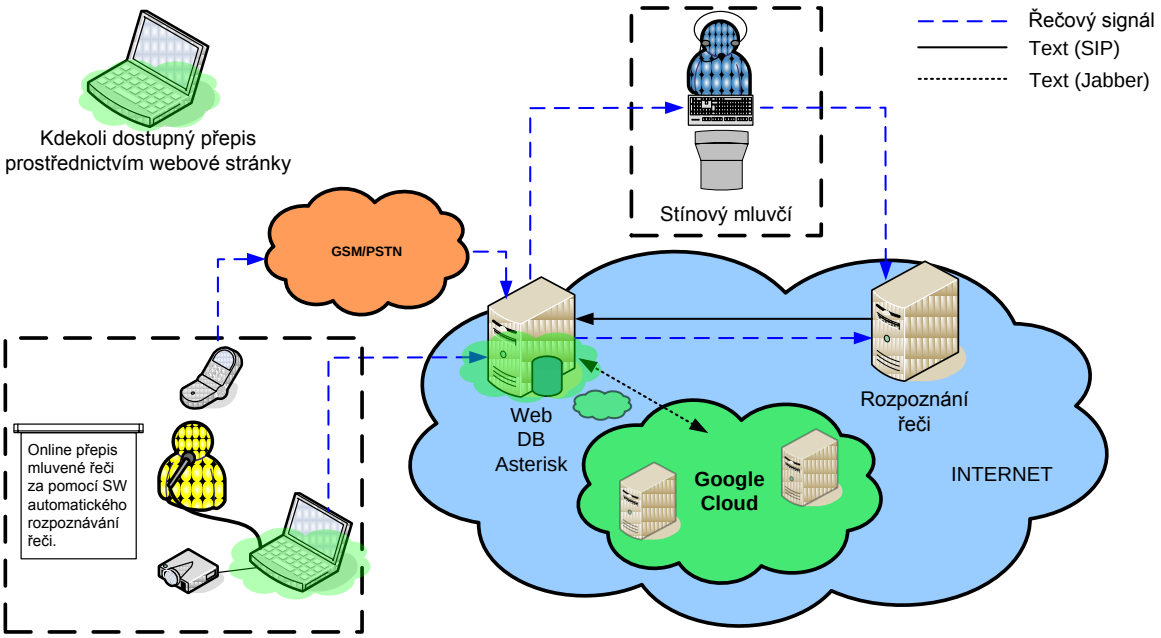
Asterisk je schopný zpracovávat různé druhy protokolů jak ve VoIP telefonii (SIP, MGCP, H.323 a IAX), tak v pevných i mobilních telefonních sítích. Velká výhoda systému spočívá právě v jeho otevřenosti a možnosti přizpůsobení se dalším různým standardům. To je důležitá vlastnost, která je využita v projektu eScribe pro propojení s prostředím “cloud computing”.

V rámci eScribe jsou využívány protokoly pro internetovou telefonii SIP (Session Initiation Protocol) a IAX (Inter-Asterisk eXchange). Pomocí těchto protokolů je možné uskutečňovat hovory zdarma odkudkoli s přístupem na Internet. Samotný hlas je pak pro potřeby přenosu v IP prostředí digitálně zakódován, většinou je použit standardní kodek G.711, jehož podpora je povinná ve všech zařízeních používající VoIP telefonii. Pro komunikaci se sítí PSTN využívá projekt eScribe propojení pomocí soukromého VoIP operátora, do mobilní sítě GSM je systém připojen pomocí GSM bran. Text vznikající rozpoznáním řeči je pak přenášen jednak využitím podpory instant messagingu v protokolu SIP a jednak protokolem XMPP/Jabber, s jehož pomocí je možné propojit infrastrukturu do Google cloudu. Technologie Googlu zatím nepodporují plně protokol SIP, proto je nutná tato protokolová konverze.

6.2 Architektura systému

Princip inovovaného systému eScribe II je zobrazen na obr. 2. Oproti původní verzi zde není zobrazena role přepisovatele z důvodu přehlednosti obrázku, avšak tato funkce zůstává samozřejmě zachována. Přepisovatel je prostřednictvím svého Google účtu připojen do systému a může psát do Google dokumentu, který mohou ostatní uživatelé zároveň prohlížet ve webovém prohlížeči. Hlas jde přepisovateli do sluchátek klasickou cestou přes ústřednu Asterisk.

Zásadní změna, která se v novém systému udála, je zavedení automatického strojového rozpoznání mluvené řeči. Rozpoznávání se děje za účasti firmy Newton Technologies a.s.[16], která se specializuje na systémy pro diktování řeči. Automatický rozpoznávač simuluje činnost přepisovatele a je do systému připojen velice podobně jako přepisovatel. Hlasový signál jde do rozpoznávače přes Asterisk ústřednu protokolem SIP.



Obr.2: Princip systému eScribe II

Na straně rozpoznávače běží SIP klient, který je standardním způsobem registrovaný na Asterisk. Tento SIP klient je ovšem speciálně upravený, aby byl schopen komunikovat s infrastrukturou rozpoznávacího serveru. Rozpoznaný text se z rozpoznávací infrastruktury vrací zpět do upraveného SIP klienta odkud je znovu prostřednictvím SIP protokolu přenesen na Asterisk. V Asterisku dochází k dalšímu zpracování zprávy, která je pro potřeby přenosu obohacena o záhlaví a je rozdělena do segmentů. Po dekódování textové informace je zpráva poslána do Google cloudu přes protokol XMPP/Jabber. Na straně cloudu je po nezbytném zpracování informace uložena do Google dokumentu, odkud je možné jeho obsah zpřístupňovat přes webové rozhraní určitým uživatelům s patřičným oprávněním pro prohlížení.

ASR systém není v současné době bezchybný a je velice citlivý na kvalitu hlasového signálu. Pro optimální (nikoliv bohužel stoprocentní) výkonnost je zapotřebí splnit mnoho požadavků jako např. přenášet hlasový signál kvalitními kodeky bez výpadků signálu na trase, používat kvalitní mikrofon a zvukové zařízení pro zpracování analogového audio signálu. Mluvčí by měl mluvit zřetelně, měl by se vystříhat přeřeknutím a jiným zvukům nenesoucí žádanou informaci. Prostředí, kde se hovoří, by mělo být prosté rušivých vlivů jako jsou další hlasy např. z publika, nebo šumy a ruchy na pozadí scény. Jisté zlepšení přináší rovněž použití určitého druhu slovníku, kdy rozpoznávací systém předpokládá, že tématicky se hovor týká např. medicínského oboru nebo může být řeč z oblasti sportu atd. V neposlední řadě se užívá techniky natrénování rozpoznávacího systému na daného mluvčího, kdy se pomocí souboru tréninkových vět rozpoznávač natrénuje na hlasový projev příslušné osoby, což opět zvyšuje úspěšnost rozpozvávání. Český jazyk je naneštěstí velice rozličný a bohatý na slovní zásobu. Rovněž není světově rozšířeným jazykem, takže problematika jeho rozpoznávání není zdaleka tak rozšířená jako např. u angličtiny. V našem systému je možné pro určité akce využít služeb tzv. stínového mluvčího, který bude poslouchat hlasový projev mluvčího a bude jej dále přemlouvat. Až řeč stínového mluvčího pak bude posílána do rozpoznávače, čímž se eliminuje řada výše zmíněných jevů negativně ovlivňující proces rozpoznávání. Úspěšnost takového rozpoznávání by pak měla být velmi vysoká.

6.3 Online zobrazování přepisu v prostředí Google cloud

Pojem cloud computing je kompletní změnou paradigmatu přístupu k řešení informatiky. Stávající model funguje na principu klient-server a prodeji licencí. Na lokálním zařízení musí být nainstalován program, kde jsou uložena i primární data. Pokud aplikace vyžaduje kooperaci, je nutné vzdáleně zavolat server. Stejně pracoval eScribe před přechodem na cloud. Uživatel musel vlastnit proprietární textový editor s tím, že data měl uložena lokálně u sebe. Pokud chtěl text ze svého počítače zobrazit vzdáleně, bylo nutné naprogramovat funkci pro odesílání a zároveň i pro zobrazování na cílovém zařízení. Tato struktura přechodem na cloud odpadá. V rámci cloud computingu jsou veškeré ICT zdroje převedeny na stranu poskytovatele (do tzv. mraku) a

Obr. 3 - Struktura modelu cloud computing

poskytovány zákazníkům formou služby. Zákazník se nemusí starat o provoz a s tím spojené náklady. Naopak mu nový model přináší mobilitu (do služeb se lze připojit odkudkoliv), jednoduchost (k provozování stačí pouze webový prohlížeč) a multiplatformnost (k připojení lze využít jakýkoliv operační systém, mobilní telefon s prohlížečem a do budoucna i hojně se rozvíjející tablety). Veškerá primární data jsou uložena na internetu, odpadá tak nutnost dvousměrné synchronizace. Uživatel se přihlašuje na vzdálené aplikace pomocí tenkého klienta, což ve většině případů představuje webový prohlížeč. K dalším výhodám cloud řešení patří škálovatelnost výkonu serverů, který se odvíjí od množství uživatelů.

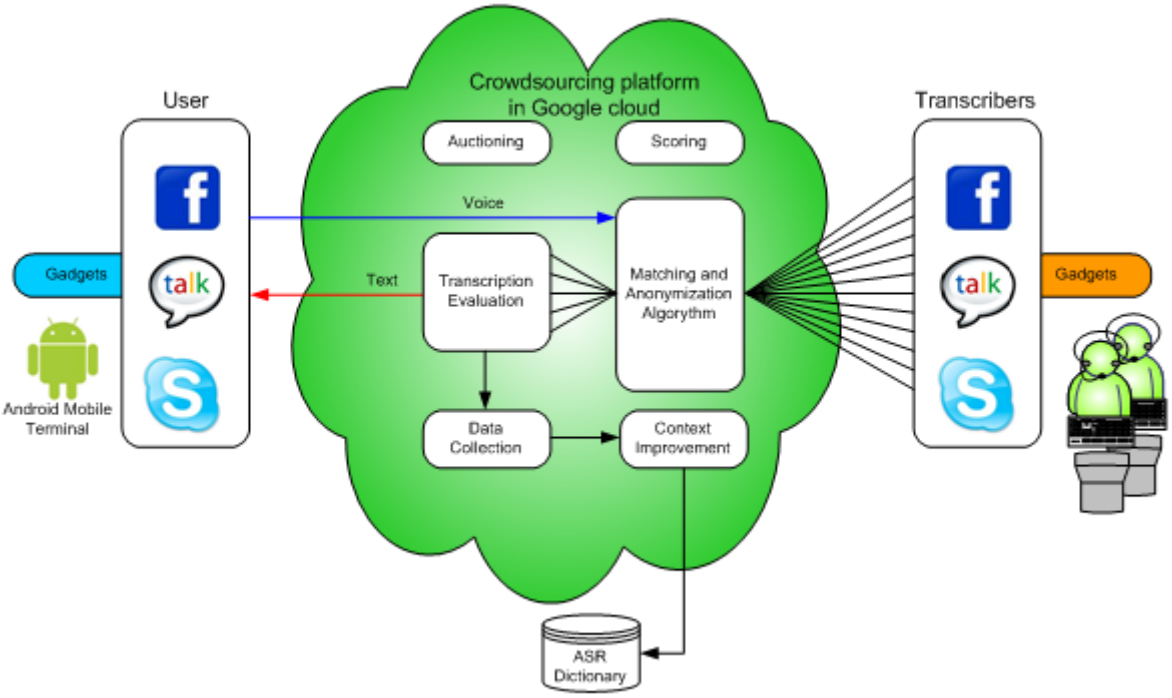
Existují tři modely cloud computingu, které se dělí podle úrovně přístupu ke službám - jde o Infrastrukturu jako službu, Platformu jako službu a Software jako službu [17].

- Infrastruktura jako služba (IAAS) je na nejnižší úrovni pomyslné pyramidy. Zákazník může konfigurovat celý systém jak potřebuje. Může tak nainstalovat operační systém a na něm budovat další aplikace. Mezi nejvýznamější providery této služby patří Amazon.
- Platforma jako služba (PAAS) je na vyšší úrovni než IAAS. Operační systém je již připraven, stejně tak prostředí pro běh programu (tzv. runtime). Mezi nejznámější patří Microsoft Azure (ASP.NET) nebo Google AppEngine (Python, Java).
- Software jako služba (SAAS) je nejběžnější formou cloud computingu. Uživatel se nestará o provoz, ale pouze se přihlašuje k již vytvořené aplikaci. Ke známějším patří Google Dokumenty, ZOHO nebo Microsoft Office 365.

Při návrhu nového řešení projektu eScribe byly využity cloud technologie firmy Google. V prvé řadě jde o Google dokumenty (SAAS) jakožto rozhraní pro psaní a skladování textů. Uživatel se přihlásí svým uživatelským jménem a heslem a může začít používat rozhraní, které připomíná běžný desktopový textový editor. Další částí systému je robot vytvořený na Google AppEngine (PAAS). Tento robot vytvořený na platformě JAVA přijímá text z rozpoznávače přes XMPP jabber protokol a posílá ho do dokumentu přes Google Document List API. Pro lepší modularitu systému budou využity tzv. gadgety, což jsou miniaplikce z nichž je možné si sestavit vlastní rozhraní dle potřeby.

7. Budoucí práce – crowdsourcing přepisovatelů v sociálních sítích

Limitujícím faktorem současných systémů přepisu řeči do textu je dostatečný počet vyškolených přepisovatelů resp. stínových mluvčí a s nimi spojené finanční náklady. Do doby než budou ASR systémy schopné spolehlivě rozpoznávat lidskou řeč i v hlučném a zarušeném prostředí, bude role lidského faktoru v obdobných projektech nezastupitelná. Velký rozvoj a popularita sociálních sítí jako např. Facebook s sebou přináší značný potenciál pro získávání přepisovatelů resp. stínových mluvčích v rámci těchto sítí. Využití lidského potenciálu pro hromadné zpracování nebo sběr „moudrosti davu“ je v anglické literatuře označován moderním pojmem crowdsoursing. Samotná myšlenka využití zkušeností velké skupiny lidí a jejich dovedností však není ničím novým a lze ji vysledovat v každém historickém období. Ale až prudký rozvoj sociálních sítí umožňuje využít tento potenciál téměř v reálném čase a zapojit obrovské množství lidí. V další etapě projektu se proto budeme zabývat crowdsourcingem přepisovatelů resp. stínových mluvčí v rámci sociálních sítí. Každý uživatel sociální sítě vybavený klávesnici, reproduktory a mikrofonem pro nás představuje možného online



Obr. 4 – Uspořádání platformy pro crowdsourcing přepisovatelů v sociálních sítích

spolupracovníka, kterého je zapotřebí oslovit, motivovat a v ideálním případě zapojit do projektu. V rámci projektu se proto budeme kromě technické realizace platformy pro crowdsourcing zabývat především motivačními nástroji, algoritmy pro vhodné přiřazení skupiny přepisovatelů jednotlivým uživatelům a v neposlední řadě anonymizačním nástrojům pro ochranu citlivých přepisů.

8. Závěr

Široce dostupný a všudypřítomný přístup ke službě přepisu řeči v reálném čase založené na moderních internetových technologiích umožní poskytovat tuto službu mnohem většímu okruhu osob s postižením sluchu. Prostřednictvím přepisu na dálku ve spolupráci s přepisovateli a s využitím systému ASR a doplněného o stínového mluvčího bude možné zajišťovat službu přepisu levněji a efektivněji než při osobní přítomnosti přepisovatele na místě konání akce a bude možné také podstatně zvětšit počet „přepisovaných“ akcí.

Díky online přepisu se neslyšícím zpřístupní kulturní, vzdělávací, společenské či jiné akce, kterých by se kvůli komunikační bariéře nemohli zúčastnit. Přepis lze využít také při individuálních jednáních neslyšících osob např. u soudu nebo na úřadě, kde je neschopnost domluvit se pro neslyšící osoby zvlášť tíživým problémem.

Námi navržená otevřená architektura asistivních hlasových služeb umožní poskytování těchto služeb nezávisle na místě, času a i typu uživatelského terminálu, protože službu je možné využívat na jakémkoliv zařízení, které umožňuje prohlížet webové stránky. Svou modulárností umožní zohlednit jednotlivé potřeby nejen neslyšících uživatelů.

Omezení projektu dané počtem přepisovatelů a stínových mluvčí zamýšlíme řešit pomocí crowdsourcingu těchto služeb v rámci sociálních sítí, což zároveň povede k rozšíření a popularizaci projektu a v neposlední řadě poskytne cenný vědecký materiál pro další výzkum a rozvoj v oblasti hlasových asistivních služeb.

9. Poděkování

Řešitelé projektu chtějí touto cestou poděkovat Nadaci Vodafone ČR [18] za její laskavou podporu a financování projektu eScribe.

Literatura

[1] Bumbalek, Z., Zelenka, J., Kencl, L.: E-Scribe: Ubiquitous Real-Time Speech Transcription for the Hearing-Impaired. 12th International Conference on Computers Helping People with Special Needs (ICCHP), 2010, Vienna, Austria.

[2] Bumbálek, Z.: Využití IP telefonie v asistivních technologiích pro neslyšící. Access server [online]. 2010, Internet: http://access.feld.cvut.cz/view.php?nazevclanku=vyuziti-ip-telefonie-v-asistivnich-technologiich-pro-neslysici&cisloclanku=2010020003

[3] Rudinsky, Mikula, Kencl, Dolezal, Garcia: Voice2Web: Architecture for Managing Voice-Application Access to Web Resources. 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services (MMNS) October 26-27, 2009, | Venice, Italy;

[4] Meggelen, J. – Madsen, L. – Smith, J.: Asterisk™: The Future of Telephony. 2nd Edition. Sebastopol: O'Reilly Media, 2007

[5] Miyoshi, S., Kuroki, H., Kawano, S., Shirasawa, M., Ishihara, Y.: Support Technique for Real-Time Captionist to Use Speech Recognition Software. In: ICCHP 2008, LNCS 5105, pp. 647–650, 2008

[6] Borodin, Y., Dausch., G., Ramakrishnan, I.V.: TeleWeb: Accessible Service for Web Browsing via Phone. In: W4A2009 collocated with WWW 2009, Madrid, Spain, April 20-21 (2009)

[7] Forman, I., Brunet, T., Luther, P., Wilson, A.: Using ASR for Transcription of Teleconferences in IM Systems.Universal Access in HCI, Part III, HCII 2009, LNCS 5616, pp. 521–529, 2009

[8] Kheir, R. and Way, T.: Inclusion of Deaf Students in Computer Science Classes Using Real-Time Speech Transcription. In Proceedings of the 12th Annual SIGCSE Conference on Innovation & Technology in computer Science Education (Dundee, Scotland, June 25-27, 2007). ITiCSE'07. ACM, New York, 192-196.

[9] The Liberated Learning Consortium, http://www.liberatedlearning.com/

[10] Wald, M.: Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time. In: ICCHP 2006, LNCS 4061, pp. 683 – 690, 2006.

[11] IBM ViaScribe, http://www-03.ibm.com/able/accessibility_services/ViaScribe-accessible.pdf

[12] Miller, M.: Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Quo, 2009

[13] http://www.escribe.cz

[14] http://www.rdc.cz

[15] Česká unie neslyšících, http://www.cun.cz, http://www.prepis.cz

[16] http://www.newtontechnologies.cz

[17] National Institute of Standards and Technology: A NIST Notional Definition of Cloud Computing, http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

[18] http://www.nadacevodafone.cz