

Systém pro automatický asistovaný překlad na mobilních zařízeních s OS Android

Bc. Václav Bárta, doc. Ing. Josef Chaloupka, Ph.D.

Abstrakt

Motivací této práce je stav CAT (computer-assisted translation) na platformě Android. Existuje velké množství strojových překladů, nikoliv však nástroj pro sofistikovaně asistovaný překlad. V současnosti nejsou tyto nástroje dostupné. Tato práce pojednává o implementaci aplikace pro asistovaný překlad pro platformu Android včetně rozpoznávání textu z obrázku. Součástí práce je seznámení s problematikou asistovaného překladu, strojového překladu a souvisejících technik. Aplikace byla naprogramována v jazyce Java a mimo jiné využívá TTS, OCR, JNI. Aplikace podporuje širokou škálu zařízení od verze SDK 8 - verze operačního systému Android 2.2.x až do Froyo.

Úvod

CAT nástroje se odlišují od běžných strojových překladů přístupem k překládanému textu. Zatímco běžný překladový nástroj nabídne pouze automatický překlad celého textu (např. aplikace ŠP překlad Google), CAT nástroje rozdělí text na menší logické části a kromě automatického překladu nabízí pokročilé vlastnosti pro zkvalitnění překladu jako tezaury, překladové paměti, terminologické databáze apod. Druhou důležitou odlišností je podpora formátů textu k překladu. CAT nástroje podporují širokou škálu formátů a tedy mimo běžný text slouží k překladu dalších typů textu jako například titulky, lokalizaci aplikací a zdrojových kódů.

Cílem této práce je navržení a implementace aplikace pro asistované překládání textu na platformě Android. Tato práce si klade za cíl vlastní implementaci CAT nástroje, nicméně další možností by mohlo být portování a úprava existující aplikace (OmegaT - Java, OpenSource) na platformu Android. OmegaT je typickým zástupcem desktopového CAT nástroje a jako takový může sloužit pro porovnání dosažené funkcionality (s určitými omezeními danými platformou).

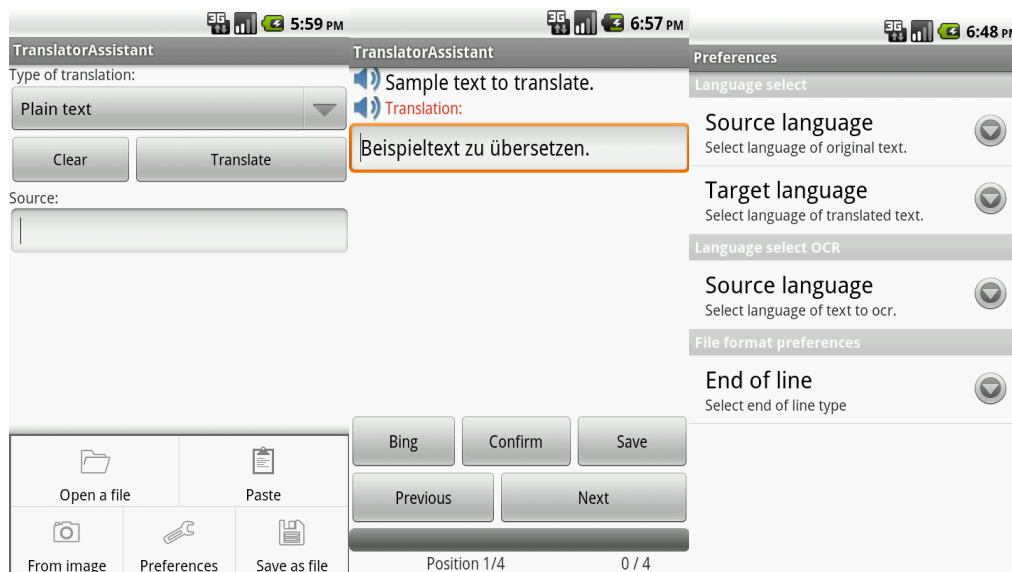
Podle zadání by měl program implementovat překlad textu po menších částech (překladových jednotkách), strojový překlad, tezaurus, podporu několika textových formátů souborů, načtení textu z obrázku (OCR textu) a TTS.

Implementace

Stavební prvky aplikace pro operační systém Android se nazývají aktivity (activity). Každá aktivita reprezentuje obrazovku aplikace, odpovídá nějaké činnosti, kterou aplikace poskytuje. Aktivita obsahuje aktivity pro zadání textu k překladu, samotný překlad, zobrazení výsledného textu, rozpoznání textu z obrázku, vyhledávání ve slovníku a nastavení aplikace. Pro GUI (**Chyba! Nenalezen zdroj odkazů.**), lokalizaci aplikace a její konfiguraci se využívají aplikační prostředky platformy Android (xml soubory).

Aplikace využívá knihovny a komponenty pro přístup k API strojového překladu (Bing) a tezaury (BigHugeThesaurus), pro detekci kódování souborů (cpdetector) a práci s nimi (andexplorer) a pro rozpoznávání textu z obrázku (Tesseract, ZoomableImageView). Rozpoznávání textu a strojový překlad podporují cca 40 jazyků, podpora TTS závisí na konkrétním zařízení, a tezaurus pouze angličtinu.

Vývoj probíhal v prostředí Eclipse (Indigo Service Release 1) s doplňkem Android Development Toolkit (16.0.1).



Obrázek 1. GUI aplikace o aktivitu pro zadání textu a typu p ekladu, p eklad a nastavení aplikace.

Výsledky rozpoznávání

Rozpoznání textu z obrazu je výpo etn náro né o doba pro rozpoznání je závislá na rozpoznávaném jazyce, na kvalitě a rozm rech rozpoznávaného obrázku a na výkonu zařízení.



Obrázek 2. P íklady OCR o z informa ní zna ky o rozpoznáno áste n o Brake Fon raoosn IT COULD SAVE YOUR LIFE, po 2x zoomu obrazu je jifl rozpoznán správný text BRAKE FOR MOOSE IT COULD SAVE YOUR LIFE o z ínského nápisu o rozpoznáno o strojový p eklad Bing: Welcome to an ancient vessel for broomcorn millet.

Záv r

Tato práce m la za cíl realizovat CAT nástroj pro opera ní systém Android. Aplikace byla realizována podle požadavk na ni kladených a svými vlastnostmi se adí mezi CAT nástroje, nicmén stále je n kolik možností jak ji dále rozvíjet. Jedná se o p ípadnou v t-í podporu formát , pouflití více slovník , možnosti ukládání vlastních p eklad k op tovnému pouflití atd.

Pod kování

Rád bych pod koval vedoucímu práce panu doc. Ing. J. Chaloupkovi, Ph.D. za cenné rady a konzultace a sle n Bc. Veronice Dufkové za jazykové korektury.

Reference

- [1] MURPHY, Mark L. *Android 2: pr vodce programováním mobilních aplikací*. Brno : Computer Press, 2011. ISBN 978-80-251-3194-7.
- [2] HEROUT, Pavel. *Java a XML*. eské Bud jovice : Kopp, 2007. ISBN 978-80-7232-307-4.