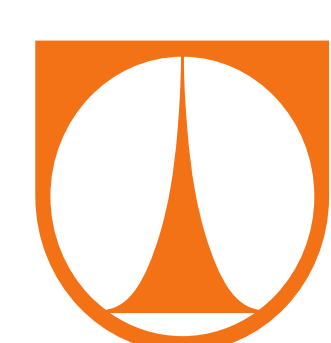


System pro automatický asistovaný překlad na mobilních zařízeních s OS Android

Bc. Václav Bárta

doc. Ing. Josef Chaloupka, Ph.D.



TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky
a mezioborových studií

Abstrakt

Motivation of this work is the state of CAT (computer-assisted translation) on the Android platform. There are many machine translators but none of them is sophisticated tool for assisted translation. This work deals with implementation of application for computer assisted translation to Android platform, even with optical character recognition. Work includes introduction into computer assisted translation, automatic translation and related techniques. Application was created by using following technologies: TTS, OCR and JNI. Application supports a wide range of devices - since SDK version 8 / 2.2.x / Froyo.

Úvod

CAT nástroje se odlišují od běžných strojových překladačů přístupem k překládanému textu. Zatímco běžný překladač nabízí pouze automatický překlad celého textu (např. aplikace „Překladač Google“), CAT nástroje rozdělí text na menší logické části a kromě automatického překladu nabízí pokročilé vlastnosti pro zkvalitnění překladu – 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 – např. titulků, lokalizací aplikací či zdrojových kódů.

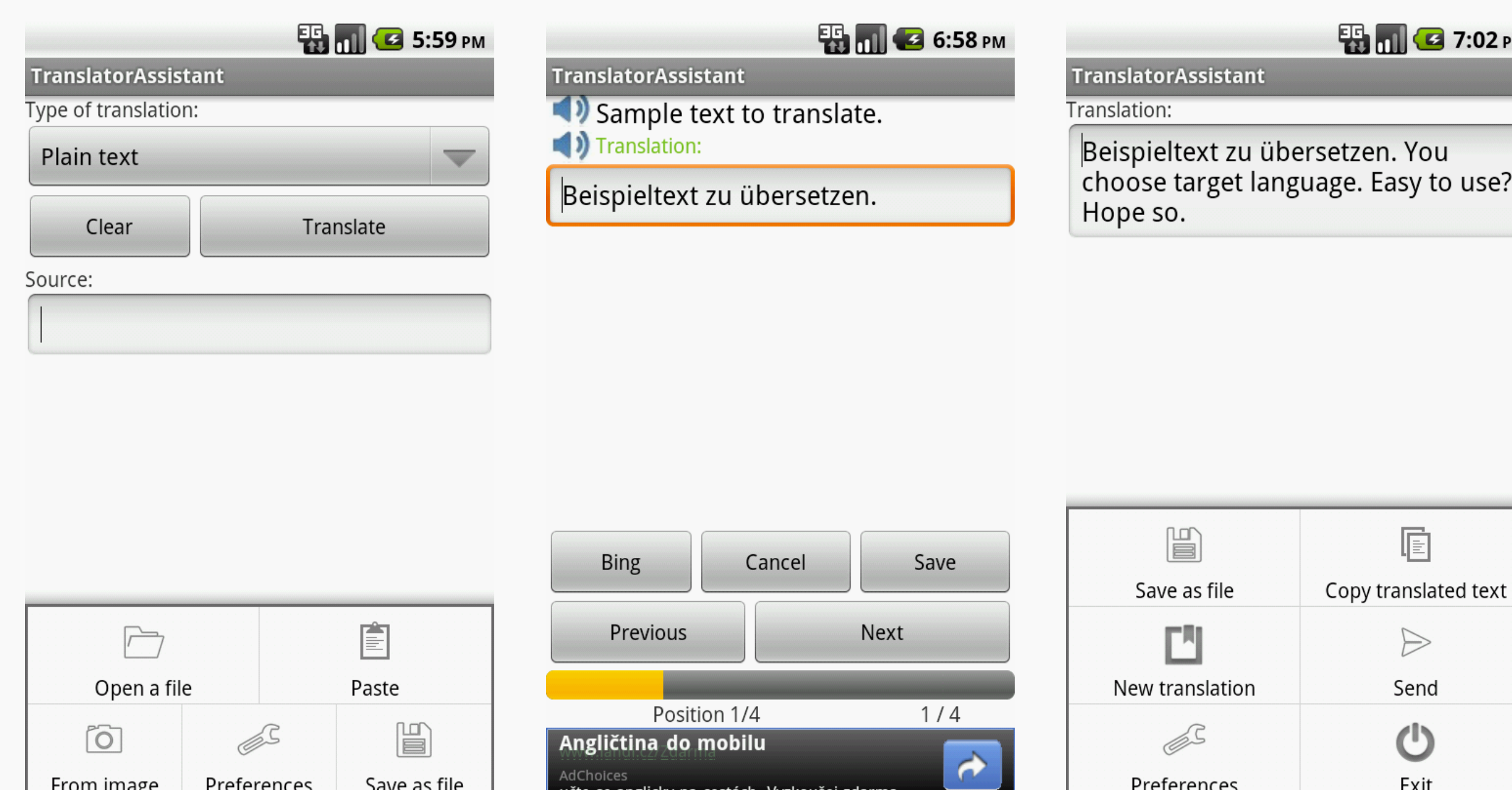
Cílem této práce je návrh a implementace aplikace pro asistované překládání textu na platformě Android. Tato práce si klade za cíl vlastní implementaci CAT nástroje, nicméně další možností by bylo portování a úprava existující aplikace (OmegaT - Java, OpenSource) na platformu Android. OmegaT je typickým 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 – aplikace obsahuje aktivity pro zadání textu k překladu, samotný překlad, zobrazení výsledného textu, rozpoznání textu z obrázku, vyhledávání ve slovníku a nastavení aplikace. Pro GUI (Obrázek 1 - GUI aplikace), lokalizaci aplikace a její konfiguraci se využívají aplikační prostředky platformy Android (xml soubory).

Aplikace využívá knihovny a komponenty pro zpřístupnění API strojového překladu (Bing) a tezauru (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.

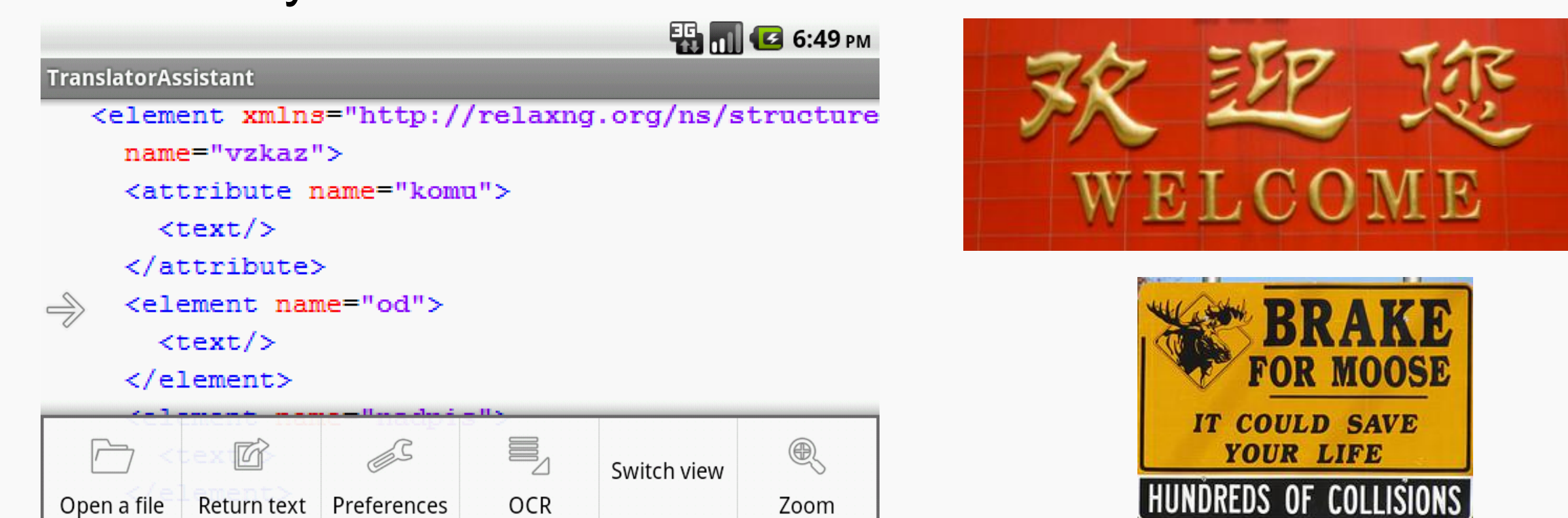
Vývoj probíhal v prostředí Eclipse (Indigo Service Release 1) s doplňkem Android Development Toolkit (16.0.1).



Obrázek 1. GUI aplikace – základní aktivity pro překlad.

Výsledky rozpoznávání

Rozpoznání textu z obrázku je výpočetně náročné – doba pro rozpoznání je závislá na rozpoznávaném jazyce, na kvalitě a rozměrech rozpoznávaného obrázku a na výkonu zařízení.



Obrázek 2. Rozpoznávání textu.

Příklady OCR

– z informační značky – rozpoznáno částečně – Brake Fon raosn IT COULD SAVE YOUR LIFE, po 2x zoomu obrázku je již rozpoznán správný text BRAKE FOR MOOSE IT COULD SAVE YOUR LIFE

– z čínského nápisu – rozpoznáno – strojový překlad Bing: Welcome to an ancient vessel for broomcorn millet.

Závěr

Tato práce měla za cíl realizovat CAT nástroj pro operační systém Android. Aplikace byla realizována podle požadavků na ni kladených a svými vlastnostmi se řadí mezi CAT nástroje, nicméně stále je několik možností jak ji dále rozvíjet. Jedná se o případnou větší podporu formátů, použití více slovníků, možnosti ukládání vlastních překladů k opětovnému použití atd.

Reference

- [1] MURPHY, Mark L. Android 2: průvodce programováním mobilních aplikací. Brno : Computer Press, 2011. ISBN 978-80-251-3194-7.
- [2] HEROUT, Pavel. Java a XML. České Budějovice : Kopp, 2007. ISBN 978-80-7232-307-4.

Kontakt

Bc. Václav Bárta - barta@mshot.eu
doc. Ing. Josef Chaloupka, Ph.D. - josef.chaloupka@tul.cz