

UTMA (Unified Translation Memory Access)

specifikace SW projektu

Úvod

V současné době tvoří lokalizace velmi nákladnou součást vývoje softwaru. Ve snaze tyto náklady snížit používají SW firmy tzv. *CAT nástroje* (Computer Assisted Translation). Hlavní součástí těchto nástrojů je tzv. *překladačová paměť* (translation memory, TM), což je databáze umožňující znovuvyužití již přeložených frází.

Je zřejmé, že čím více v TM obsažených frází překládaný text obsahuje, tím levnější je překlad. Nedostatkem této technologie je, že autora textu nijak nenutí používat fráze v TM již obsažené. Přístup, který usnadňuje autorovi používání terminologie konzistentní s TM, se nazývá *Translation Oriented Authoring*. Zatímco překladačových pamětí existuje velké množství, systémů pro Translation Oriented Authoring není mnoho.

Cíl

Cílem našeho projektu je vytvořit několik modulů umožňujících a usnadňujících Translation Oriented Authoring v rámci IDE (popřípadě i v jiném SW, ve kterém vzniká překládaný obsah) a propojit tyto moduly s již existujícími překladačovými pamětmi pomocí univerzálního rozhraní.

Toto propojení umožní autorům překládaného obsahu dotazovat se překladačové paměti na existující termíny a zajistit tak vyšší konzistenci použité terminologie.

Spolupráce

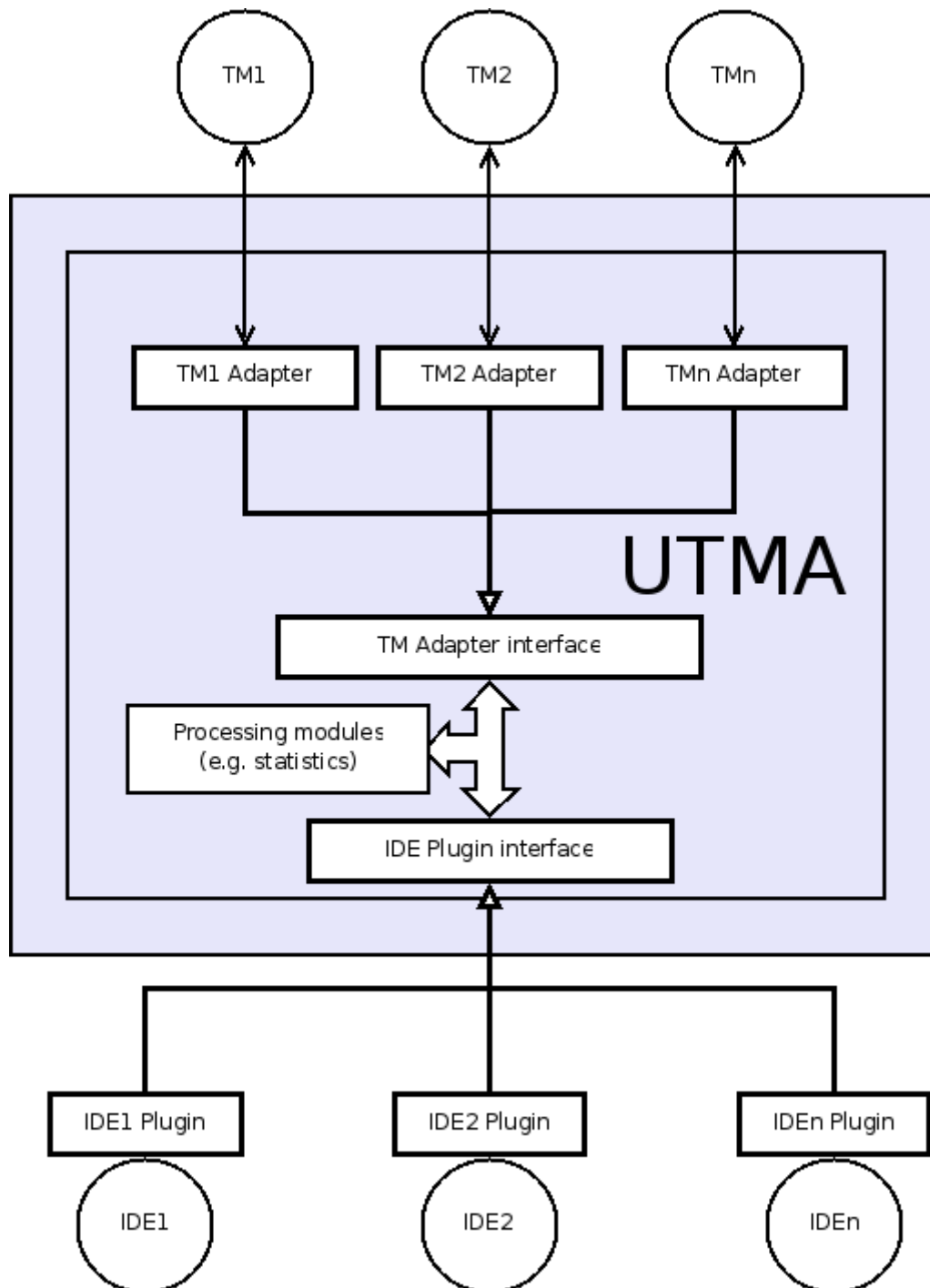
Práce na projektu bude probíhat ve spolupráci s firmou Sun Microsystems (Prague Globalization). Její zástupci poskytnou řešitelskému týmu informace potřebné pro analýzu problému.

Požadavky

Návrh architektury i výběr technologií ovlivňují především následující požadavky:

- platformní nezávislost
- snadná rozšiřitelnost

Architektura



Systém bude mít několik částí:

- Nejnižší vrstvu tvoří *TM Adaptery*, moduly pro zapojení překladové paměti do UTMA systému, abstrahující rozdíly mezi rozhraními jednotlivých překladových pamětí a implementující jednotné rozhraní využívané vyššími vrstvami systému UTMA.
- Na opačné straně stojí rozhraní *IDE Plugin*, které využijí pluginy z vývojových prostředí či textových procesorů, aby se mohly připojit k UTMA systému.

- Několik modulů pro různá IDE, které budou k překladové paměti přistupovat přes jednotné rozhraní IDE Plugin.
- „Processing modules“ mohou být využity např. pro logování dotazů, což se může hodit k tvorbě řady zajímavých statistik (např. „kolik dotazů přišlo na termíny v němčině“).

Rozsah

Vytvoříme TM adaptér pro alespoň jednu překladovou paměť (konkrétně pro Idiom WorldServer). Není jisté, zda bude možné vytvořit TM adaptéry pro jiné překladové paměti. Většina překladových pamětí totiž nemá jasně definované a zdokumentované přístupové aplikační rozhraní.

Součástí bude několik modulů pro různá IDE: modul pro NetBeans a jedno až dvě jiná vývojová prostředí. Moduly budou integrovány do IDE: pro fráze, které uživatel ve zdrojovém kódu použije, umožní modul najít alternativní fráze spolu s kritérii, podle kterých se uživatel bude moci rozhodnout, kterou frází použít (tímto kritériem může být například počet jazyků, do nichž existuje překlad dané fráze). Výběr alternativní fráze provede buď sám uživatel na základě svého uvážení, nebo modul automaticky. Podoba pluginů bude závislá na hostitelském IDE, hlavně na tom, jaké API dané IDE poskytuje autorům pluginů. Mimo tyto moduly vytvoříme ještě webové rozhraní, které (nezávisle na jakémkoliv IDE) bude poskytovat stejné služby jako pluginy (tedy pro danou frází najde alternativní fráze spolu s kritérii).

Jádro systému UTMA bude sloužit jako server pro pluginy. Taktéž překladové paměti mohou mít serverový charakter. Z toho důvodu bude systém navržen tak, aby jádro systému, moduly pro IDE a překladová paměť mohly běžet na různých počítačích.

Technologie

- Java
- Idiom WorldServer

Vedoucí

Mgr. Pavel Ježek, Katedra SW inženýrství, MFF UK

Členové

Michal Kebrt, Pavel Římský, Petr Štěpán, Lenka Trochtová