

**Program děkana Fakulty informatiky MU pro podporu studentských
výzkumných a vývojových projektů**

Návrh projektu

Identifikační kód projektu¹

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | U | N | I | 3 | 3 | / | 0 | 3 | 2 | 0 | 0 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | |
|--|--|
| Základní údaje | |
| Garant projektu ² (jméno, příjmení, email) | doc. RNDr. Petr Sojka, Ph.D., sojka@fi.muni.cz |
| Vedoucí projektu ³ (jméno, příjmení, email) | |
| Název projektu česky | Rozšíření redakčního systému odborného časopisu s podporou exportu do digitální knihovny |
| Název projektu anglicky | Extensions to the Specialist Journal Processing System with the Capability of Exporting to a Digital Library |
| WWW stránka projektu | http://www.fi.muni.cz/lemma/math-journal/ |
| Doba trvání projektu v měsících | 8 |

Složení studentského řešitelského týmu (uveďte jméno, příjmení, e-mail a UČO všech studentů)

Michal Růžička, mruzicka@mail.muni.cz, 143424

Anotace projektu česky (4–5 řádků, tato informace bude zveřejněna na www stránkách FI)

Cílem projektu je rozšířit komplexní, na typografickém systému TeX založený, redakční systém pro vydávání odborného časopisu o podporu MathML, přidat nezávislý modul pro validaci výstupů na straně digitální knihovny a zvýšit přístupnost výstupů systému.

¹ O číslo je třeba požádat pracovníka oddělení VaV FI MU (např. e-mailem).

² Garantem projektu musí být zaměstnanec FI. Studenti doktorského studia zaměstnanci zpravidla nejsou. Pokud má být projekt veden doktorandem, je třeba ho uvést jako vedoucího projektu a nalézt formálního garanta z řad zaměstnanců FI (např. školitele dotyčného doktoranda).

³ Vedoucím projektu je buď garant, nebo jím pověřený student doktorského studia. Vyplňujte jen v případě, kdy garant není současně vedoucím projektu.

Anotace projektu anglicky (4–5 řádků, tato informace bude zveřejněna na www stránkách FI)

The goal of the project is to extend the comprehensive specialist journal processing system based on the TeX typesetting system to support MathML, add an independent module intended for digital-library-side metadata validation and improve accessibility of outputs of the system.

Cíle projektu

Stručně popište

1. současný stav projektu (podporu mohou získat pouze projekty, které již nějaké výsledky vykazují),
 2. cíle projektu, způsob jejich dosažení, předpokládané výstupy, forma zveřejnění, licence vyvíjených programů⁴,
 3. časový harmonogram projektu (déle běžící projekty je třeba rozdělit na etapy, jejichž délka nepřesahuje tři měsíce; u každé etapy je třeba uvést **konkrétní a kontrolovatelné** výstupy, jejichž plnění bude kontrolovat garant).
1. Komplexní, na typografickém systému TeX založený, redakční systém pro vydávání odborného časopisu je (v návaznosti na projekt úspěšně řešený v rámci prvního kola Programu děkana FI MU pro podporu studentských výzkumných a vývojových projektů) od začátku roku 2008 produkčně nasazen pro vydávání matematického časopisu Archivum Mathematicum [1] a jeho metadata jsou importována do České digitální matematické knihovny (DML-CZ) [2]. Další čísla časopisu jsou v přípravě.
- O projektu byl prezentován článek [3] v rámci DML 2008 Workshop (Birmingham, UK, 27 July 2008) [4], který byl přidružen k mezinárodní konferenci Mathematical Knowledge Management 2008 (Birmingham, UK, 28–30 July 2008) [5].
2. Úkolem systému je maximálně automatizovat přípravu vydávání jednotlivých čísel časopisů pro ulehčení práce redakce, a zároveň připravovat kvalitně označovaná metadata pro potřeby digitálních knihoven. Hlavním cílem projektu je další rozvoj výše uvedeného redakčního systému s cílem obohatit jej o využívání nových moderních technologií. Jedná se zejména o:
- Obohacení XML metadat exportovaných pro potřeby digitálních knihoven o možnost využití jazyka MathML pro uchování matematických výrazů obsažených v metadatach.
 - Implementace samostatného modulu pro provoz na straně digitální knihovny, který bude provádět automatizovanou validaci výstupů (zejména metadat) produkovaných redakčním systémem před jejich importem do digitální knihovny.
 - Průzkum množností (a kde to bude možné, tak i implementace) postupů pro zvýšení přístupnosti systémem produkovaných výstupů. Jednat by se mohlo např. o využívání pokročilých vlastností, které poskytuje formát PDF (vytváření tagovaných PDF nesoucích informaci o struktuře dokumentu, vytváření PDF v normovaných formátech, např. PDF/A určeném pro dlouhodobou archivaci PDF dokumentů a normalizovaném jako ISO 19005-1:2005 apod.).
- Vedle toho bude v rámci projektu zapracována další funkcionalita, jejíž potřeba vyvstane ze strany redakcí / digitálních knihoven během řešení projektu (potencionálně by se mohlo jednat např. o zajištění původu a integrity obsahu digitálním podpisem přidávaným k elektronickým verzím časopisů / metadatům apod.).

⁴ Podporu je možno získat pouze pro vývoj *svobodného software*, zveřejněného např. pod GPL nebo LGPL licencí, nikoliv pro vývoj komerčních produktů.

Redakční systém je založen na volně dostupných nástrojích, především pak na typografickém systému TeX s makronadstavbou LaTeX a vychází z redakčního systému projektu CEDRAM [6]. Pro přípravu metadat je použit open-source nástroj Tralics [7].

Pro generování PDF výstupů se používá program pdfTeX – rozšířená implementace původního systému TeX. Možnost generování PDF dokumentů využívajících výše uvedených pokročilých vlastností z prostředí systému TeX je v současné době u programu pdfTeX ve fázi vývoje v podobě nových nízkoúrovňových primitivů jazyka. V tomto ohledu tedy do značné míry bude cílem projektu tuto podporu testovat, a pokud to již bude možné, tak ji do redakčního systému implementovat v podobě vysokoúrovňových maker.

Výstupem projektu bude nová verze redakčního systému obohacená o export metadat pro digitální knihovnu se zápisem matematiky v XML jazyce MathML a samostatný modul pro provoz na straně digitální knihovny pro validaci výstupů redakčního systému před jejich importem do knihovny. Pokud to bude možné, systém bude rozšířen o využívání technik vedoucích ke zlepšení přístupnosti výstupů generovaných systémem. Zdrojové texty systému budou zveřejněny na domovské stránce projektu. Všechny části systému jsou licencovány pod některou ze svobodných licencí (konkrétní typ licence se u jednotlivých použitých nástrojů liší). Zázemí projektu poskytnou laboratoře NLP [8] a LEMMA [9].

3. Projekt bude zahájen v září 2009 a předpokládaná doba trvání projektu je osm měsíců. V první tříměsíční etapě dojde k implementaci rozšíření exportu XML metadat pro digitální knihovnu o podporu exportu matematických výrazů v XML jazyce MathML. Dále dojde k zahájení prací na implementaci samostatného modulu pro validaci výstupů redakčního systému před importem do digitální knihovny.

Ve druhé tříměsíční etapě projektu dojde k implementaci případných rozšíření systému vyžádaných ze strany redakcí / digitálních knihoven. Dále bude dokončena implementace samostatného modulu pro validaci výstupů systému a započnou práce na vylepšení přístupnosti výstupů redakčního systému.

V závěrečné dvouměsíční etapě projektu dojde k dokončení všech prací z předchozích etap. Dokončena bude zejména implementace vylepšení vedoucích ke zlepšení přístupnosti výstupů generovaných redakčním systémem.

Odkazy:

- [1] Elektronická edice časopisu Archivum Mathematicum. <http://www.emis.de/journals/AM/>
- [2] Domovská stránka projektu DML-CZ. <http://dml.muni.cz/>
- [3] Růžička, Michal. Automated Processing of TeX-Typeset Articles for a Digital Library. In *DML 2008: Towards Digital Mathematics Library*. First edition. Brno : Masaryk University, 2008. s. 167–176. ISBN 978-80-210-4658-0.
- [4] Domovská stránka DML 2008 workshop. <http://www.fi.muni.cz/~sojka/dml-2008.xhtml>
- [5] Domovská stránka Mathematical Knowledge Management 2008.
- [6] Domovská stránka projektu CEDRAM. <http://www.cedram.org/>
- [7] Domovská stránka programu Tralics. <http://www-sop.inria.fr/apics/tralics/>
- [8] Domovská stránka laboratoře NLP. <http://nlp.fi.muni.cz/>
- [9] Domovská stránka laboratoře LEMMA. <http://www.fi.muni.cz/lemma/>

| Finanční rozvaha (pouze neinvestiční prostředky, uvádět v tisících Kč na jedno desetinné místo) | | | |
|--|--|--------------|--------------------------|
| | položka | návrh | rozhodnutí komise |
| 1 | Stipendia pro členy řešitelského týmu ⁵ | 25,6 | |
| 2 | Cestovné ⁶ | 0,0 | |
| 3 | Literatura, materiál, drobný majetek ⁷ | 0,0 | |
| 4 | Celkem | 25,6 | |

| Zdůvodnění finanční rozvahy⁸ |
|---|
| <p>Plánována je výplata měsíčních stipendií ve výši 3 200 Kč po dobu 8 měsíců řešení projektu. Náročnost jednotlivých prací, zejména v případě druhé a třetí etapy (tedy při přípravě využití vývojových verzí pdfTeXu pro podporu generování PDF s využitím pokročilých vlastností formátu) je obtížné odhadnout. Předpokládá se, že se bude jednat řádově o stovky pracovních hodin.</p> <p>V době běhu projektu není plánována participace jeho řešitele jako spolupracovníka na jiném grantovém projektu. Práce na tomto projektu v žádném případě nebudou financovány z jiných zdrojů.</p> |

Souhlas garanta

Navrhovaný projekt je v souladu s obecnými cíli *Programu pro podporu studentských výzkumných a vývojových projektů* popsanych v článku 1 Pokynu děkana Fakulty informatiky MU 2/2008. Cíle projektu považuji za reálné a požadované finanční prostředky za přiměřené. Souhlasím s odbornou garancí projektu a přejímám zodpovědnost za kontrolu plnění stanovených cílů projektu. O případných problémech budu neprodleně informovat proděkana pro výzkum a vývoj.

Datum:

podpis garanta

⁵ Maximální výše stipendia je 3 200 Kč měsíčně pro jednoho studenta. Doktorští studenti mohou být členy řešitelského týmu, ale nelze pro ně požadovat stipendium.

⁶ Cestovné lze požadovat pouze v dobře odůvodněných případech, kdy jsou pracovní cesty nevyhnutelnou podmínkou úspěšného naplnění cílů projektu.

⁷ Vlastníkem pořízené literatury a drobného hmotného majetku je Fakulta informatiky MU.

⁸ Explicitně uveďte, zda (a případně kteří) členové řešitelského týmu participují jako spolupracovníci na nějakém jiném grantovém projektu.