



Projekt LISp-Miner

<http://lispminer.vse.cz>

J. Rauch, M. Šimůnek



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

Obsah

List of tasks in the selected group

Bad Loans

Kvalitní půjčka

Kvalitní půjčka (1)

Rozložení pohledů

Špatná půjčka

Test 0 Succ.

- Systém LISp-Miner
- Vývoj systému v dlouhém období
- Zajímavosti implementace
- Distribuovaný výpočet



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected

Bad Loans

Kvalitní půjčka

Kvalitní půjčka (1)

Rozložení pohledů

Špatná půjčka

Test 0 Success

System LISp-Miner

- Metoda GUHA (od roku 1966)
- Předchozí implementace (IBM 370, PC-GUHA...)
- <http://lispminer.vse.cz>
- Akademické prostředí
- Cíle
 - použití ve výuce
 - použití pro výzkum v oblasti DM
 - použití pro řešení reálných DM úloh
- Modulární a vrstvená architektura
- Důraz na rychlost výpočtu
 - generování a verifikace co nejrychlejší (i pro složité zadání úloh)
- Vývoj od roku 1996
 - cca 1 mil. programových řádků
 - vývojové prostředí MSVC++



LM Database: LM Banka DB

Datamining database: LM Banka Data

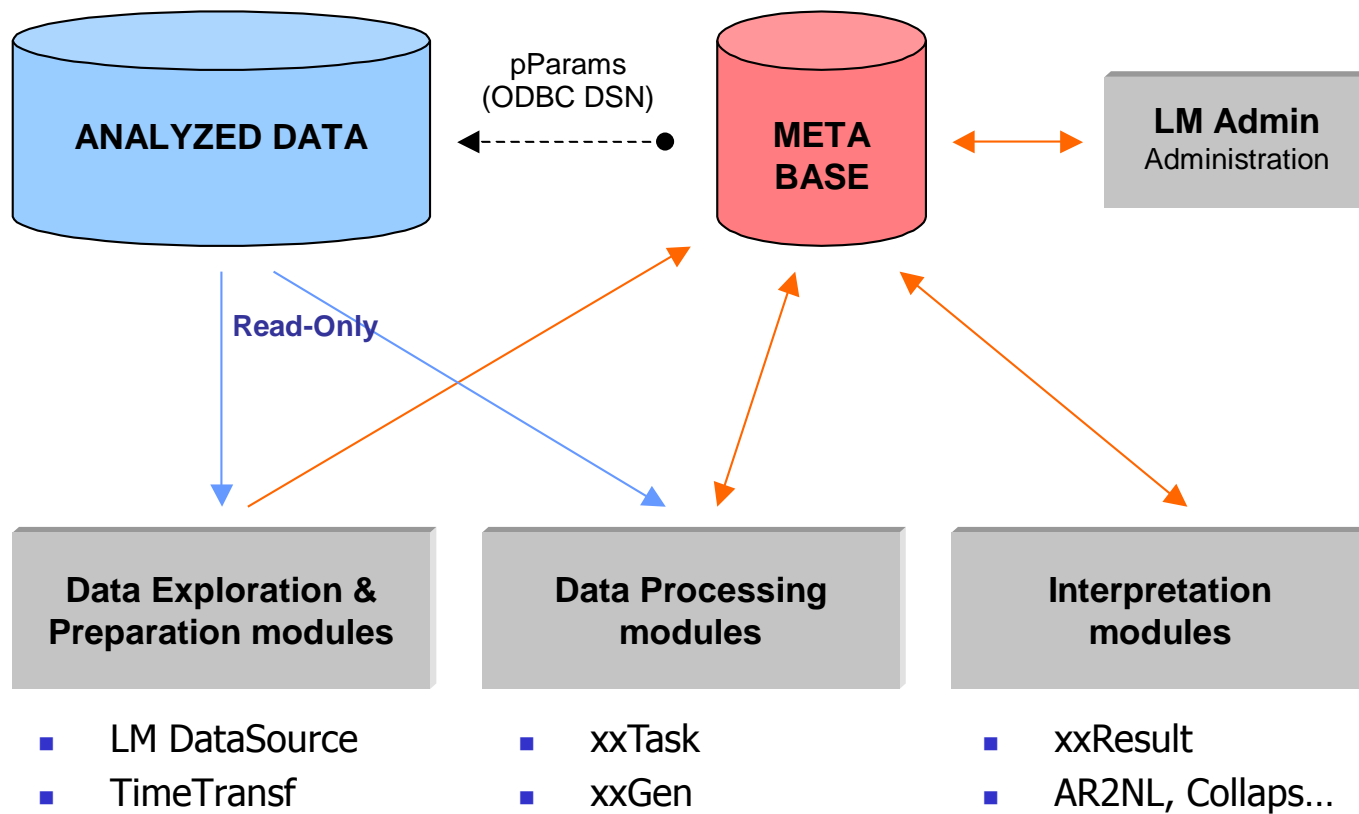
Selected task group

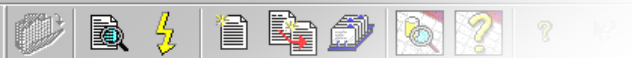
Default task-group

List of tasks in the selected group

- Bad Loans
- Kvalitní půjčka
- Kvalitní půjčka (1)
- Pozložení pohledávky
- Špatná půjčka
- Test 0 Succ...

Architektura systému





LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected

Bad Loans
Kvalitní půjčka
Kvalitní půjčka (1)
Rozložení pohledů
Špatná půjčka
Test 0 Succ.

Zajímavosti z implementace

- Bitové řetězce
- Procházení variant cedentů
- Distribuovaný výpočet pomocí gridu



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected

Bad Loans
Kvalitní půjčka
Kvalitní půjčka (1)
Rozložení pohledů
Špatná půjčka
Test 0 Succ...

Bitové řetězce

- Výpočet frekvencí do čtyřpolní tabulky
- Binární reprezentace vstupních dat
 - Karta kategorie
 - Př. Veličina „Město“ a kategorie Praha, Brno, Ostrava
 - Karta_Kategorie_Praha
 - 000110100011100011100000111100001...
 - Karta literálu
 - logická disjunkce karet kategorií
 - Karta cedentu
 - logická konjunkce karet literálů
- Bitové operace nad kartami
 - AND, OR, NOT
- Zjištění počtu záznamů v kartě



Zjištění počtu záznamů v kartě

- Bitový řetězec

- 000110100011100011100000111100001...

- Úkol

- spočítat počet „1“ v bitovém řetězci

- Možnosti

- přes index pole

- přes bitový posun a „zero-flag“

- předpočítaná tabulka



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected group

Bad Loans
Kvalitní půjčka
Kvalitní půjčka (1)
Rozložení pohledů
Špatná půjčka
Test 0 Success

Procházení variant cedentů

- **BOOL PrepareNextCedent ()**
 - sekvenční procházení všech možných kombinací literálů a jejich koeficientů v cedentu
 - výhody a nevýhody sekvenčního přístupu (stejně jako u cursorů a streamů)
- **Stavová tabulka**
 - přehledný popis velmi složitého algoritmu
 - snadná změna implementace (zejména při ladění)
- **Přeskoky na základě frekvencí**
 - vynechání částí stavového prostoru
 - velmi zrychlují verifikaci



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected

Bad Loans

Kvalitní půjčka

Kvalitní půjčka

Rozložení po

Špatná půjčka

Test 0 Suč

Stavová tabulka (část)

Stav	Délka	Frekvence	Lze prodloužit	Je další literál	Lze výměna veličiny	Lze zkrátit	Akce
1	0	**	**	**	**	**	ProdlouzeniCedentu
2	1	AA	AA	**	**	**	ProdlouzeniCedentu
3	1	AA	NN	AA	**	**	DalsiLiteral
4	1	AA	NN	NN	AA	**	VymenaVeliciny
5	1	AA	NN	NN	NN	AA	ZkraceniCedentu
6	1	AA	NN	NN	NN	NN	KonecVetveVBNC
7	1	NN	**	AA	**	**	DalsiLiteral
8	1	NN	**	NN	AA	**	VymenaVeliciny
9	1	NN	**	NN	NN	AA	ZkraceniCedentu
10	1	NN	**	NN	NN	NN	KonecVetveVBNC
11	2	**	**	AA	**	**	DalsiLiteral
12	2	**	**	NN	AA	**	VymenaVeliciny
13	2	**	**	NN	NN	AA	ZkraceniCedentu
14	2	**	**	NN	NN	NN	KonecVetveVBNC



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected group

Bad Loans
Kvalitní půjčka
Kvalitní půjčka (1)
Rozložení pohledů
Špatná půjčka
Test 0 Success

Využití gridu

- Některé úlohy opravdu velmi složité
 - generování a verifikace až stovek milionů relevantních otázek (zejména nové procedury jako Action4ft-Miner)
- Rozdělení úlohy na menší části a rozdělení na více počítačů
 - PC-Grid na Technical University of Tampere
 - obyčejné počítače na učebnách, u sekretárek apod., které jsou zaregistrovány u Grid-Serveru, že mohou při nízkém zatížení sloužit pro externě zadaný výpočet
 - framework řeší otázky vhodného přidělení výpočtu, bezpečnosti...
- Kdy má smysl optimalizovat?
- Režie
 - rozdělení úlohy na nezávislé části
 - přenos dat
 - agregace výsledků



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

List of tasks in the selected group

Bad Loans
Kvalitní půjčka
Kvalitní půjčka (1)
Rozložení pohledů
Špatná půjčka
Test 0 Succ...

Metoda Monte-Carlo

- Ideální úloha pro distribuovaný výpočet
 - málo vstupních dat, která nutné distribuovat mezi klienty
 - snadné rozdělení
 - každý „pokus“ je nezávislý a čím více pokusů, tím lépe
 - snadná agregace výsledků
 - zjištění průměrné hodnoty za všechny pokusy
- Výpočet hodnoty Pi



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

Implementace

List of tasks in the selected

Bad Loans

Kvalitní půjčka

Kvalitní půjčka (1)

Pozložení pohledávky

Špatná půjčka

Test 0 Success

- Podařilo se naimplementovat do
 - 4ft-Mineru a Ac4f-Mineru
 - transparentní z pohledu uživatele
- Velmi dobré výsledky
- Dobrá škálovatelnost přidáním další PC do gridu



Nr.	Task	Local Time	Grid Time	Grid	Active nodes
1	A (4ft-Miner)	0h 14m 14s	0h 9m 48s	TUT	30
2	B (4ft-Miner)	3h 48m	0h 21m 38s	TUT	30
3	B (4ft-Miner)	3h 48m	0h 53m 01s	UEP	4
4	B (4ft-Miner)	3h 48m	1h 5m 28s	UEP	5
5	C (Act4ft-Miner)	30h 23m	1h 06m 35s	TUT	24
6	C (Act4ft-Miner)	30h 23m	5h 44m 57s	UEP	5



LM Database: LM Banka DB

Datamining database: LM Banka Data

Selected task group

Default task-group

Závěr

List of tasks in the selected

Bad Loans

Kvalitní půjčka

Kvalitní půjčka (1)

Rozložení pohledů

Špatná půjčka

Test 0 Succ.

- Systém LISp-Miner
- Další vývoj

<http://lispminer.vse.cz>