

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

DATAMINING Z JABBERU

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

JAROSLAV SENDLER

BRNO 2010



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

DATAMINING Z JABBERU

DATAMINING FROM JABBER

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

JAROSLAV SENDLER

VEDOUcí PRÁCE

SUPERVISOR

JOZEF MLÍCH SUPERVISOR.TITLE.P

BRNO 2010

Abstrakt

Výtah (abstrakt) práce v českém jazyce.

Abstract

Výtah (abstrakt) práce v anglickém jazyce.

Klíčová slova

Klíčová slova v českém jazyce.

Keywords

Klíčová slova v anglickém jazyce.

Citace

Jaroslav Sendler: Datamining z jabberu, bakalářská práce, Brno, FIT VUT v Brně, 2010

Datamining z jabberu

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana ...

.....

Jaroslav Sendler

14. prosince 2010

Poděkování

Zde je možné uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc.

© Jaroslav Sendler, 2010.

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Kapitola 1

Úvod

- 1.1 Musíme mít co říci
- 1.2 Musíme vědět, komu to chceme říci
- 1.3 Musíme si dokonale promyslet obsah
- 1.4 Musíme psát strukturovaně?

Kapitola 2

XMPP

Pro usnadnění a lepší pochopení budou v následující kapitole rozebrány základní stavební kameny protokolu Extensible Messaging and Presence Protocol (XMPP). Samotný protokol je datován do roku 2004 (březen), kdy na něj byl pojmenován jabber. Původní projekt jabber byl vytvořen roku 1998 autorem Jeremie Miller, jež ho založil na popud nesvobodných uzavřených IM služeb. Měl mít tři základní vlastnosti - jednoduchost a srozumitelnost pro implementaci, jednoduše rozšiřitelný a otevřený. Základní vlastnosti a výhody klientů a serverů budou popsány níže. Roku 1999, 4. ledna vytvořil první server se jménem jabber. Komunita vývoje se chopila iniciativy a napsala klienty pro různé platformy (Linux, Macintosh, Windows), kteří dokázali se serverem komunikovat. Roku... byl přidán mezi RFC (request of comments - žádost o komentáře) dokumenty. Základní normy jsou RFC 3920 (obecná specifikace protokolu) a RFC 3921 (samotný instant messaging a obrazy stavu). Další zdokumentovaná rozšíření jsou vydávána v podobě tzv. XEP (XMPP Extension Protocol) dokumentů, starším jménem JEP (Jabber Enhancement Proposal). Dnešní počet těchto norem se blíží k číslu 300. Každý XEP obsahuje status, stav vývoje (schválení), ve kterém se nachází. XMPP protokol je postaven na obecném značkovacím jazyce XML, proto vlastnosti popsané v kapitole platí i pro tento protokol.

1. ==znaková sada

2.1 Jabber

Dnes známý jako komunikační platforma založená na protokolu XMPP. Vzdáleně jej lze přirovnat k dnes již na slávu upadajícímu software ICQ (I Seek You) využívající protokol OSCAR (Open System Communication in Realtime - otevřený systém pro komunikaci v reálném čase).

2.2 XML

Jazyk XML (eXtensible Markup Language), metajazyk pro deklaraci strukturovaných dat, je jádrem protokolu XMPP. Samotný jazyk vznikl rozšířením metajazyka SGML, jež slouží pro deklaraci různých typů dokumentů. Základní vlastností je jednoduchá definice vlastních značek (tagů). Dokument XML se skládá z elementů, jež mezi sebou navzájem zanořovávají. Vyznačíme je pomocí značek - počáteční a ukončovací.

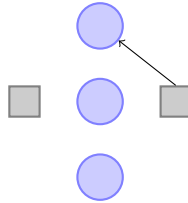
1. obrazek struktura XML stream...

Základní jednotkou komunikace je stanza. Obsahuje 3 elementy message, presence a iq, je? ka?dý má sv?j jednozna?ný význam.

2.3 Message

2.4 IQ

2.5 presence



Klient

+—+—+

Server

obrazek distribuovana architektura

```

+—+—+
— Jabber —
— client — +—+—+
+—+—+ — —
— — Jabber — +.....+
+—+—+ — server — : :
— — : :
+—+—+ : :
— : Internet :
+—+—+ : : —+
: : —
: : +—+—+
: : — —
: : — Jabber —
+.....: — server —+
— — —
+—+—+ +—+—+
— Jabber —
— client —

```




2.6 Knihovny

Jabber je realizován jako otevřený XML standart pro instant messaging formát, proto existuje mnoho programovacích jazyků, kterým je práce s ním usnadněna pomocí knihovny. Mezi nejznámější patří C (iksemel, libstrophe, Loudmouth), C++ (gloox, Iris), JAVA (JabberBeans, Smack, JSO, Feridian, Emite, minijingle), .NET (Jabber-Net, agsXMPP SDK), Python (JabberPy, PyXMPP, SleekXMPP, Twisted Words), Ruby (XMPP4R, Jabber4R, Jabber::Simple, Jabber::Bot), Perl (Net-Jabber) některé níže budou lehce rozebrány a vyzviženy jejich hlavní přednosti.

iksemel

JabberBeans

Jabber-Net

JabberPy

Kapitola 3

Dataming

3.1 Metody dolovani dat

3.2 Dolovani znalosti z databazi

Kapitola 4

Developer

4.1 Cile

4.2 Slepa ulicka

4.3 knihivna, jazyk

4.4 Jine produkty

Kapitola 5

Nikdy to nebude naprosto dokonalé

Kapitola 6

Typografické a jazykové zásady

6.1 Co to je normovaná stránka?

Kapitola 7

Záv?r

Příloha A

Struktura XMPP

IQ

```
<iq>ELEMENT
  to
  from
  id
  type
    ='get'
    ='set'
    ='result'
    ='error'
  NAMESPACE
```

Tabulka A.1: Část algoritmu *Non maximal supression*

PRESENCE

```
<presence>ELEMENT
  to
  from
  id
  type
    ='available'
    ='unavailable'
    ='probe'
    ='subscribe'
    ='unsubscribe'
    ='subscribed'
    ='unsubscribed'
    ='error'
  <show/>
    away
    chat
    dnd
    normal
    xa
  <status/>
  <priority/>
  <error/>
NAMESPACE
```

Tabulka A.2: Část algoritmu *Non maximal supression*

MESSAGE

```
<message>ELEMENT
  to
  from
  id
  type
    = '..'
    = 'chat'
    = 'groupchat'
    = 'headline'
    = 'error'
  <body/>
    Offline
    Delivered
    Displayed
    Composing
  <subject/>
  <thread/>
  <error/>
  <xi/>
NAMESPACE
```

Tabulka A.3: Část algoritmu *Non maximal supression*

Příloha B

Slovník pojmu

Stanza

Roster

Disco

VCard