

Spejd Web Service

Szymon Acedański

Specyfikacja projektu

8. Implement a Spejd web service.

Czy już zaliczyłem? ;)

```
from SimpleXMLRPCServer import SimpleXMLRPCServer

# Create server
server = SimpleXMLRPCServer(("localhost", 8000))

def whoami():
    return "Spejd, a shallow parser Web Service!"

# Register whoami() function.
server.register_function(whoami)

# Run the server's main loop
server.serve_forever()
```


Co to jest Spejd?

Co to jest Spejd?

- Parser powierzchniowy stworzony w IPI PAN

Co to jest Spejd?

- Parser powierzchniowy stworzony w IPI PAN
- Rozpowszechniany jako Open Source

Co to jest Spejd?

- Parser powierzchniowy stworzony w IPI PAN
- Rozpowszechniany jako Open Source
- Reguły – kaskada gramatyk regularnych

Co to jest Spejd?

- Parser powierzchniowy stworzony w IPI PAN
- Rozpowszechniany jako Open Source
- Reguły – kaskada gramatyk regularnych
- Dezambiguacja

Co to jest Spejd?

- Parser powierzchniowy stworzony w IPI PAN
- Rozpowszechniany jako Open Source
- Reguły – kaskada gramatyk regularnych
- Dezambiguacja
- Parsowanie powierzchniowe (grupy składniowe)

Co to jest Web Service?

Co to jest Web Service?

Web Service

Strona internetowa

Co to jest Web Service?

Web Service

- Usługa, której nie widać.

Strona internetowa

- HTML, CSS, te sprawy...

Co to jest Web Service?

Web Service

- Usługa, której nie widać.
- Zestaw funkcji lub komponentów dostępnych przez sieć przy użyciu ustalonego protokołu, np. XML.

Strona internetowa

- HTML, CSS, te sprawy...
- Czasem bywa interfejsem użytkownika dla web service. W końcu ludzie nie mówią w XMLu.

Co to jest Web Service?

Web Service

- Usługa, której nie widać.
- Zestaw funkcji lub komponentów dostępnych przez sieć przy użyciu ustalonego protokołu, np. XML.
- Jednolity sposób komunikacji, możliwość podpięcia różnych klientów.

Strona internetowa

- HTML, CSS, te sprawy...
- Czasem bywa interfejsem użytkownika dla web service. W końcu ludzie nie mówią w XMLu.
- Bywa jednym z klientów.

Wykorzystanie projektu

Wykorzystanie projektu

- Edukacja: zajęcia dla studentów, szkolenia

Wykorzystanie projektu

- Edukacja: zajęcia dla studentów, szkolenia
- Doraźne eksperymenty, testowanie reguł tworzonych ręcznie bądź automatycznie

Wykorzystanie projektu

- Edukacja: zajęcia dla studentów, szkolenia
- Doraźne eksperymenty, testowanie reguł tworzonych ręcznie bądź automatycznie
- Wsparcie dla badań

Wykorzystanie projektu

- Edukacja: zajęcia dla studentów, szkolenia
- Doraźne eksperymenty, testowanie reguł tworzonych ręcznie bądź automatycznie
- Wsparcie dla badań
- Możliwość wykorzystania usługi na potrzeby innych serwisów
 - w końcu po to robimy Web Service, żeby nie tylko klikać przez przeglądarkę

Dotychczasowy „rynek”

Web [Show options...](#) Results 1 - 10 of about 64,100 for [shallow parser online](#). (0.20 seconds)

[Cognitive Computation Group - EOH Demonstration](#)  
Shallow parsing is the process of identifying syntactical phrases, such as noun phrases, ... [
[See an online demonstration](#) | [Back to the Main Page](#)]
l2r.cs.uiuc.edu/~cogcomp/eoh/chunker.html - 4k - [Cached](#) - [Similar pages](#) - 

[PDF] [A language-independent shallow-parser Compiler](#)  
File Format: PDF/Adobe Acrobat - [View as HTML](#)
demo is available **online** on the author's homepage. 3.2 A **Shallow-Parser** for English. We wanted to evaluate our compiler on more ...
www ldc.upenn.edu/acl/P/P01/P01-1043.pdf - [Similar pages](#) - 
by A Kinyon - [Cited by 15](#) - [Related articles](#) - [All 15 versions](#)

[CiteULike: Skeletons in the parser: Using a shallow parser to ...](#)  
Register and you can start organising your references **online**. ... We describe a simple approach for integrating **shallow** and deep **parsing**. ...
www.citeulike.org/user/kvorg/article/1476088 - 38k - [Cached](#) - [Similar pages](#) - 

[CiteULike: A language-independent shallow-parser compiler](#)  
15 Mar 2007 ... CiteULike is a free **online** bibliography manager. ... NJ, USA EP - 337 TI - A language-independent **shallow-parser** compiler PB - Association ...
www.citeulike.org/user/yoavg/article/1165413 - 32k - [Cached](#) - [Similar pages](#) - 
[More results from www.citeulike.org »](#)

[jmlr-announce: JMLR special issue on shallow parsing is now available](#)  
18 Mar 2002 ... is pleased to announce the Special Issue on Machine Learning Approaches to **Shallow Parsing**, available **online** at <http://www.jmlr.org>. ...
www.ph.tn.tudelft.nl/PRInfo/journals/msg00253.html - 5k - [Cached](#) - [Similar pages](#) - 

Dotychnczasowy „rynek”

Software error:

Can't connect to 3497: Reason: Connection refused

For help, please send mail to the webmaster (you@example.com), giving this error message and the time and date of the error.

Wymagania funkcjonalne

Wymagania funkcjonalne

- Wybór reguł Spejda
 - z listy predefiniowanych zestawów
 - lub przez podanie własnego zestawu

Wymagania funkcjonalne

- Wybór reguł Spejda
 - z listy predefiniowanych zestawów
 - lub przez podanie własnego zestawu
- Możliwość podania tekstu do analizy
 - w formacie IPI PAN
 - lub jako czystego tekstu

Wymagania funkcjonalne

- Wybór reguł Spejda
 - z listy predefiniowanych zestawów
 - lub przez podanie własnego zestawu
- Możliwość podania tekstu do analizy
 - w formacie IPI PAN
 - lub jako czystego tekstu
- Dostęp do wyników
 - wizualizacja na stronie
 - oraz możliwość pobrania jako XML

... w języku API

"Simplicity is the ultimate sophistication"

Leonardo da Vinci

... w języku API

"Simplicity is the ultimate sophistication"

Leonardo da Vinci

```
String parse(Enum rulesMode,  
             Enum textMode,  
             String rules,  
             String text)
```

```
rulesMode: PREDEFINED, CUSTOM  
textMode: PREDEFINED, PLAIN, XML
```


... w języku API

"Simplicity is the ultimate sophistication"

Leonardo da Vinci

No dobrze, dobrze. Zrobię tych kilka dodatkowych funkcji, dzięki którym na przykład będzie można pobrać predefiniowane gramatyki oraz teksty. :)

Wizualizacja wyniku

W jaki sposób dobrze pokazać wynik działania parsera powierzchniowego?

Wizualizacja wyniku

The screenshot shows a web browser window with the URL `http://korpus.pl/poliqarp/poliqarp.php`. The page is titled "Wyszukiwanie" (Searching) and contains a search form. The search query is `[pos~adj] [base~noga & case~~gen]` with the word "w" entered in the search box. The search results are displayed under the "WYNIKI" (Results) tab. The query is repeated as "Zapytanie: [pos~adj] [base~noga & case~~gen]". The results show 31 findings, with the first 10 displayed. A button "Następne 10" (Next 10) is visible. The results are presented in a table-like format with three columns: the original text snippet, the morphological analysis of the word, and the word's meaning or context.

Text Snippet	Morphological Analysis	Meaning/Context
jeszcze raz, jedna z	żelaznych [żelazny:adj:pl:gen:f:pos] nóg [noga:subst:pl:gen:f]	łóżka sypnęła iskrami. Zaraz
nawet gdyby	dziewczęcych [dziewczęcy:adj:pl:gen:f:pos] nóg [noga:subst:pl:gen:f]	, raczej nie zaryzykuje utraty
Niżycki wypatrzył biel	damskich [damski:adj:pl:gen:f:pos] nóg [noga:subst:pl:gen:f]	. - Co dalej?
zamierzam ich używać do łapania	gołych [goły:adj:pl:gen:f:pos] nóg [noga:subst:pl:gen:f]	oparto na wysokim, przyniesionym

Wizualizacja wyniku

Part of Speech Tagger Example

The boy went to see his lovely father.

Lexicon: ☒ Early Modern English
☐ Nineteenth Century Fiction

Options: ☐ Show continuous word numbers
☐ Show standardized spelling
☐ Show lemma

S#	W#	Word	Tag
1	1	The	dt
1	2	boy	n1
1	3	went	vvd
1	4	to	pc-acp
1	5	see	vvi
1	6	his	po31
1	7	lovely	j
1	8	father	n1
1	9	.	.

Wizualizacja wyniku

Grammar:

```
{<DT>?<NN>?<JJ>*<NN.*>+}
```

Prev Grammar Next Grammar

Development Set (2/100)

Chancellor/NNP of/IN the/DT Exchequer/NNP Nigel/NNP Lawson/NNP 's/POS
restated/VBN commitment/NN to/TO a/DT firm/NN monetary/JJ policy/NN has/VBZ
helped/VBN to/TO prevent/VB a/DT freefall/NN in/IN sterling/NN over/IN the/DT
past/JJ week/NN ./.

Wizualizacja wyniku

Kotek	subst:sg:nom:m2
może	fin:sg:ter:imperf
skończyć	inf:perf
na	prep:loc
dnie	subst:sg:loc:f
rzeki	subst:sg:gen:f
!	interp

Wizualizacja wyniku



Kotek	subst:sg:nom:m2
może	fin:sg:ter:imp:perf
skończyć	inf:perf
na	prep:loc
dnie	subst:sg:loc:f
rzeki	subst:sg:gen:f
!	interp

Wizualizacja wyniku

Wizualizacja wyniku

Kotek	może	skończyć	na	dnie	rzeki	!
subst:sg: :nom:m2	fin:sg: :ter:imperf	inf:perf	prep:loc	subst:sg: :loc:f	subst:sg: :gen:f	interp

Wizualizacja wyniku

Kotek	może	skończyć	na	dnie	rzeki	!
subst:sg: :nom:m2	fin:sg: :ter:imperf	inf:perf	prep:loc	subst:sg: :loc:f	subst:sg: :gen:f	interp

Kotek	subst:sg:nom:m2	może	fin:sg:ter:imperf	skończyć	inf:perf		
na	prep:loc	dnie	subst:sg:loc:f	rzeki	subst:sg:gen:f	!	interp

Wizualizacja wyniku



Wizualizacja wyniku



[NP Kotek subst:sg:nom:m2] może fin:sg:ter:imperf skończyć inf:perf
[PP na prep:loc [NP dzień subst:sg:loc:f [NP rzeki subst:sg:gen:f]]] ! interp

Wizualizacja wyniku

NP

może

skończyć

PP

!

Kotek

fin:sg:
:ter:imperf

inf:perf

na

NP

interp

subst:sg:
:nom:m2

prep:loc

dnie

NP

subst:sg:
:loc:f

rzeki

subst:sg:
:gen:f

Pozostałe wymagania

Pozostałe wymagania

- Do uruchomienia na Linuxie przy użyciu narzędzi Open Source

Pozostałe wymagania

- Do uruchomienia na Linuxie przy użyciu narzędzi Open Source
- Dostępność w dwóch językach: polskim i angielskim

Technologie

Technologie

- Java
 - ponieważ Spejd jest napisany w Javie

Technologie

- Java
 - ponieważ Spejd jest napisany w Javie
- Java RMI
 - “Java Remote Method Invocation (Java RMI) provides a **simple and direct** model for distributed computation with Java objects”

Technologie

- Java
 - ponieważ Spejd jest napisany w Javie
- Java RMI
 - “Java Remote Method Invocation (Java RMI) provides a **simple and direct** model for distributed computation with Java objects”
- Google Web Toolkit
 - ponieważ jest mi znany :)

Pomysły na przyszłość

Pomysły na przyszłość

- Użycie Morfeusza jako analizatora morfosyntaktycznego dla wpisywanych tekstów.

Pomysły na przyszłość

- Użycie Morfeusza jako analizatora morfosyntaktycznego dla wpisywanych tekstów.
- Automatyczna ewaluacja gramatyki przy danym korpusie wzorcowym.

Pomysły na przyszłość

- Użycie Morfeusza jako analizatora morfosyntaktycznego dla wpisywanych tekstów.
- Automatyczna ewaluacja gramatyki przy danym korpusie wzorcowym.
- Parsowanie fragmentów korpusu IPI PAN, które można wybrać podając zapytanie w języku Morfeusza.

Pomysły na przyszłość

- Użycie Morfeusza jako analizatora morfosyntaktycznego dla wpisywanych tekstów.
- Automatyczna ewaluacja gramatyki przy danym korpusie wzorcowym.
- Parsowanie fragmentów korpusu IPI PAN, które można wybrać podając zapytanie w języku Morfeusza.

Dziękuję za uwagę!