Understanding the Hidden Web

Pierre Senellart

Journées GEMO — 2nd June 2005
The Hidden Web

Definition (Hidden Web)

The set of webpages (which may or may not be dynamically generated) not accessible from the hyperlinked structure of the World Wide Web.

Size estimate (2001) : 500 times larger than the surface Web.

How to understand it and benefit from its content?
The Hidden Web

Definition (Hidden Web)

The set of webpages (which may or may not be dynamically generated) not accessible from the hyperlinked structure of the World Wide Web.

Size estimate (2001) : 500 times larger than the surface Web.

How to understand it and benefit from its content?
The Hidden Web

Definition (Hidden Web)

The set of webpages (which may or may not be dynamically generated) not accessible from the hyperlinked structure of the World Wide Web.

Size estimate (2001) : 500 times larger than the surface Web.

How to understand it and benefit from its content?
Web Service Semantic Interpretation Process

World Wide Web
Web Service Semantic Interpretation Process

World Wide Web → discovery → HTML form

- UDDI
- WSDL
Web Service Semantic Interpretation Process

World Wide Web \(\xrightarrow{\text{discovery}}\) HTML form
- UDDI
- WSDL
\(\xrightarrow{\text{wrappers}}\) WSDL
Web Service Semantic Interpretation Process

World Wide Web → discovery → HTML form

World Wide Web → discovery → WSDL

Analyzed Web Services → analysis → WSDL
Web Service Semantic Interpretation Process

- World Wide Web
  - discovery
  - HTML form
    - UDDI
    - WSDL
  - wrappers

- Analyzed Web Services
  - analysis
  - WSDL
  - indexing

- Web Services Index
Understanding the Hidden Web

Pierre Senellart

Introduction

Process description

Discovery
Wrappers
Semantic Analysis
Indexing and Querying

Summary

Web Service Semantic Interpretation Process

World Wide Web

\(\text{discovery}\)

HTML form

\(\text{WSDL}\)

\(\text{UDDI}\)

wrappers

Analyzed Web Services

\(\text{analysis}\)

\(\text{WSDL}\)

Web Services Index

\(\text{indexing}\)

\(\text{analyzing}\)

Queries

Result
Crawling the World Wide Web for:

- HTML forms implementing a Web Service
- UDDI registries
- WSDL descriptions
- Other resources (XML, HTML, Web as a full-text index...)

Only interested in Web Services with no side effects:

<table>
<thead>
<tr>
<th>Ok</th>
<th>Not Ok</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow Pages</td>
<td>Booking services</td>
</tr>
<tr>
<td>Publication databases</td>
<td>Mailing list management</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Web Service Discovery

Crawling the World Wide Web for:
- HTML forms implementing a Web Service
- UDDI registries
- WSDL descriptions
- Other resources (XML, HTML, Web as a full-text index...)

Only interested in Web Services with no side effects:

Ok
- Yellow Pages
- Publication databases
- ...

Not Ok
- Booking services
- Mailing list management
- ...
Crawling the World Wide Web for:
- HTML forms implementing a Web Service
- UDDI registries
- WSDL descriptions
- Other resources (XML, HTML, Web as a full-text index...)

Only interested in Web Services with no side effects:
- Ok
  - Yellow Pages
  - Publication databases
- Not Ok
  - Booking services
  - Mailing list management
  - ...
Web Service Discovery

Crawling the World Wide Web for:
- HTML forms implementing a Web Service
- UDDI registries
- WSDL descriptions
- Other resources (XML, HTML, Web as a full-text index...)

Only interested in Web Services with **no side effects**:

**Ok**
- Yellow Pages
- Publication databases
- ...

**Not Ok**
- Booking services
- Mailing list management
- ...
Analyzing the structure of:

- HTML forms
  ![HTML Form Example](image)

- Result webpages
  ![Result Webpage Example](image)
Conceptual Model

IsA ontology of concepts (simple DAG)

- Person
  - Man
  - Woman
- Publication
  - Proceedings
  - Article
  - Book

$n$-ary typed roles

- AuthorOf(Person,Publication)
- HasName(Person,Name)
Conceptual Model

- **IsA ontology of concepts** (simple DAG)

  ![Diagram]

  - Person
    - Man
    - Woman
  - Publication
    - Proceedings
    - Article
    - Book

- **n-ary typed roles**
  - AuthorOf(Person, Publication)
  - HasName(Person, Name)
Services and queries

Example

**Service giving authors from publication titles**

\[ A^* \leftarrow \text{WrittenBy}(P,A), \text{HasTitle}(P,T), \text{Input}(T) \]

Query

Service with no input

Example

\[ <A,T^*>^* \leftarrow \text{WrittenBy}(P,A), \text{Article}(P), \text{HasTitle}(P,T), \text{KeywordOf}("xml",P) \]
Services and queries

Example

**Service** giving authors from publication titles

\[ A^* \leftarrow WrittenBy(P,A), HasTitle(P,T), Input(T) \]

Query

Service with no input

Example

\[ <A,T^*>^* \leftarrow WrittenBy(P,A), Article(P), HasTitle(P,T), KeywordOf("xml",P) \]
Semantic Interpretation of a Service

How to analyze a Web Service into this formalism?

- Field labels and variable names
- Example requests
- Concrete type descriptions
- Linguistic analysis of plain text descriptions
Semantic Interpretation of a Service

How to analyze a Web Service into this formalism?

- Field labels and variable names
- Example requests
- Concrete type descriptions
- Linguistic analysis of plain text descriptions
Semantic Interpretation of a Service

How to analyze a Web Service into this formalism?

- Field labels and variable names
- Example requests
- Concrete type descriptions
- Linguistic analysis of plain text descriptions
Semantic Interpretation of a Service

How to analyze a Web Service into this formalism?

- Field labels and variable names
- Example requests
- Concrete type descriptions
- Linguistic analysis of plain text descriptions
Web Service Indexing and Querying

Given a query, represented as an Analyzed Web Service, how to know which known web services to query?

Issues:

- Subsumption of input/output parameters
- Missing input parameters
- Composition of webservices
Given a query, represented as an Analyzed Web Service, how to know which known web services to query?

Issues:

- **Subsumption** of input/output parameters
- Missing input parameters
- Composition of webservices
Given a query, represented as an Analyzed Web Service, how to know which known web services to query?

Issues:

- Subsumption of input/output parameters
- Missing input parameters
- Composition of webservices
Given a query, represented as an Analyzed Web Service, how to know which known web services to query?

Issues:

- **Subsumption** of input/output parameters
- Missing input parameters
- Composition of webservices
Web Service Semantic Interpretation Process

Understanding the Hidden Web
Pierre Senellart

Introduction
Process description
Discovery
Wrappers
Semantic Analysis
Indexing and Querying
Summary

World Wide Web

HTML form
WSDL
UDDI
discovery
wrappers

Analyzed Web Services

WSDL
analysis

Web Services Index

indexing

Queries

querying

Results