

Understanding the Hidden Web

Pierre Senellart



Journées GEMO — 2nd June 2005

The Hidden Web

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart



World Wide Web

Introduction

Process
description

Discovery

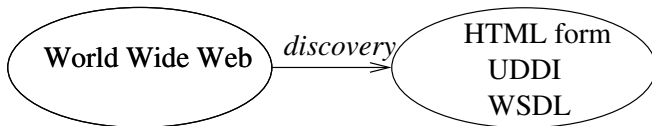
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Semantic Interpretation Process



Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

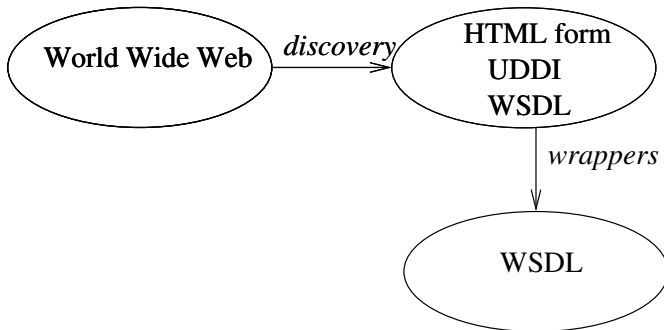
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Semantic Interpretation Process



Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

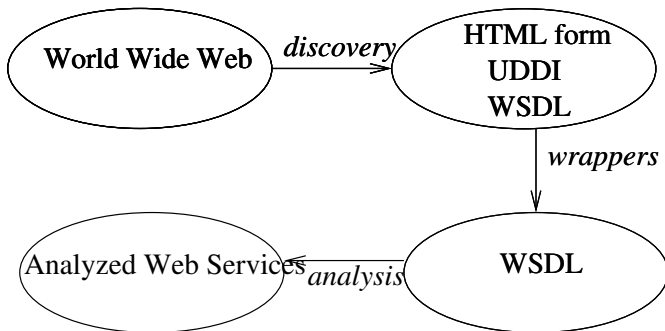
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Semantic Interpretation Process



Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

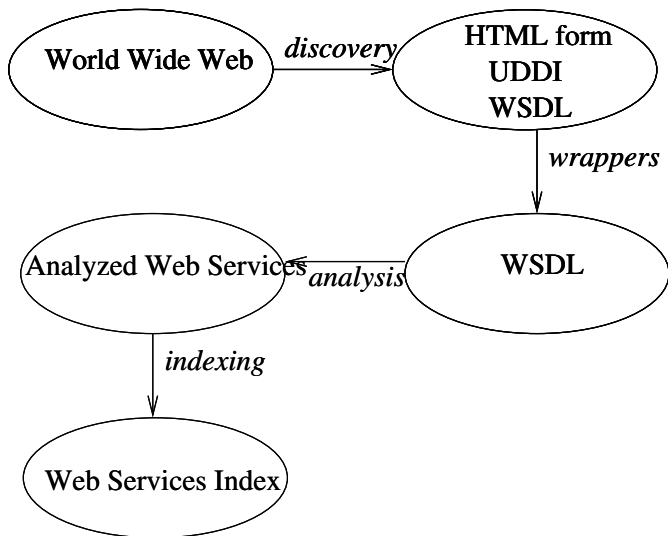
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Semantic Interpretation Process



Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

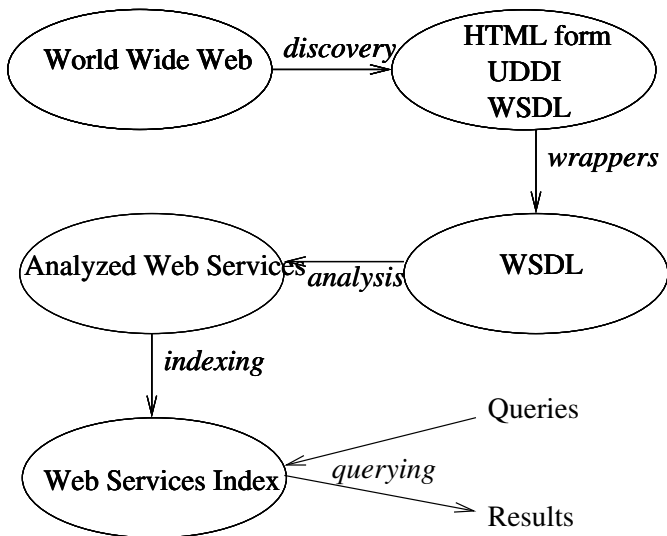
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Semantic Interpretation Process



Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

Web Service Discovery

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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
- ...

Wrapping Web Service Descriptions

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers


Semantic Analysis

Indexing and
Querying

Summary

Analyzing the **structure** of:

- HTML forms

 Inria	<input type="text"/> Nom	Saisir un nom de Projet ou sélectionner dans cette liste. <ul style="list-style-type: none">A3ACACIAACESADEPTADMIN LIPADMLORADMREN
	<input type="text"/> Prénom	
	<input type="text"/>	
	<input type="text"/>	
Rechercher Réinitialisation		

- Result webpages

Recherche sur le site de l'INRIA

Interrogation sur tout le site Inria

Vous pouvez exclure les rapports de recherche et les rapports d'activité de votre recherche :

[Aide](#) 


110 documents répondent aux critères de recherche.

1. [Serge ABITEBOUL](#)

 Serge ABITEBOUL. Scientific leader of the VERSO project. Rocquencourt research unit. 01 39 63 52 65 (Secretary office :) Homepage. webmaster@inria.fr.

URL: http://www.inria.fr/personnel/Serge_Abiteboul_en_html
dernière modification: 05 Nov 02 - taille: 865 - Unknown.

2. [Serge ABITEBOUL](#)

 Serge ABITEBOUL. Responsable scientifique du projet VERSO. Unité de recherche de Rocquencourt. 01 39 63 52 65 (Secrétariat :) Page personnelle...

URL: http://www.inria.fr/personnel/Serge_Abiteboul_fr.html
dernière modification: 05 Nov 02 - taille: 878 - Français

Conceptual Model

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

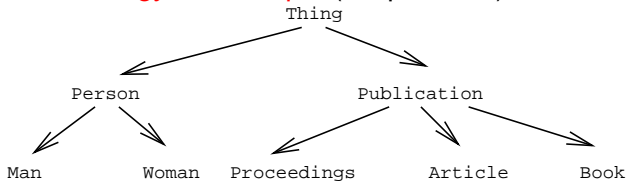
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

- IsA ontology of concepts (simple DAG)



- n -ary typed roles
 - AuthorOf (Person, Publication)
 - HasName (Person, Name)

Conceptual Model

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

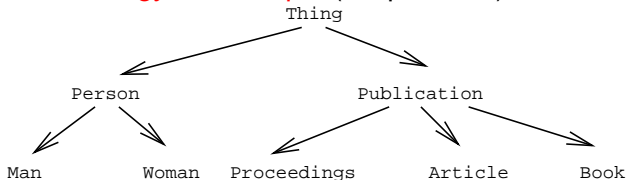
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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



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

Services and queries

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

$$\langle A, T^* \rangle^* \leftarrow \text{WrittenBy}(P,A), \text{Article}(P), \text{HasTitle}(P,T), \text{KeywordOf}(\text{"xml"}, P)$$

Services and queries

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

$$\langle A, T^* \rangle^* \leftarrow \text{WrittenBy}(P,A), \text{Article}(P), \text{HasTitle}(P,T), \text{KeywordOf}(\text{"xml"}, P)$$

Semantic Interpretation of a Service

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery
Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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 Indexing and Querying

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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 Indexing and Querying

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

Summary

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 Indexing and Querying

Understanding
the Hidden
Web

Pierre
Senellart

Introduction

Process
description

Discovery

Wrappers

Semantic Analysis

Indexing and
Querying

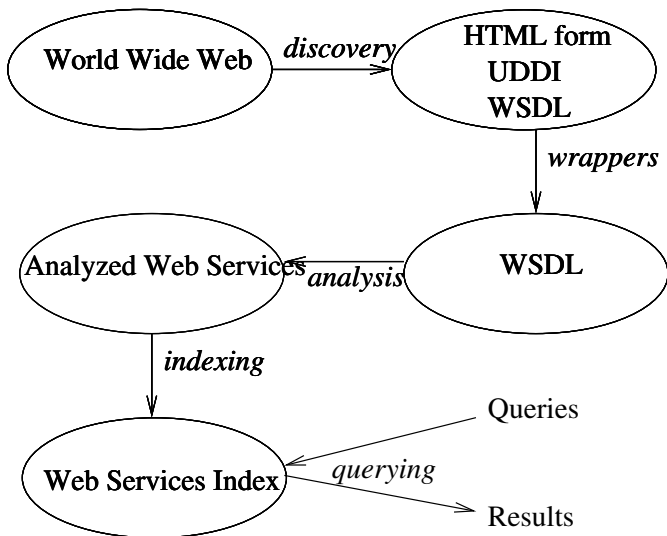
Summary

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