

**ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC MÁY TÍNH**



BÁO CÁO SEMINAR MÔN HỌC

TÍNH TOÁN LƯỚI

Đề tài:

**SEMANTIC GRID: FROM CONCEPTS TO
IMPLEMENTATION**

GVGD : TS. Phạm Trần Vũ

Học viên : Nguyễn Văn Nhất

Phạm Thị Bích Liên

Nguyễn Duy Thiện

TP. HCM, 06 / 2011

Mục lục

I.	GRID NGỮ NGHĨA	3
1.	Giới thiệu	3
2.	Web ngữ nghĩa.....	4
2.1	Ontology	5
2.2	Kiến trúc của web ngữ nghĩa	6
2.2.1	Unicode	6
2.2.2	URI.....	6
2.2.3	XML.....	6
2.2.4	XML Schema	7
2.2.5	RDF.....	7
2.2.6	RDF Schema	8
2.2.7	OWL	10
2.2.8	Logic	10
2.2.9	Proof.....	10
2.2.10	Trust.....	10
3	Grid ngữ nghĩa	10
3.1	Khái niệm.....	10
3.2	Ba lớp khái niệm trong hạ tầng tính toán	11
3.3	Các yêu cầu và thử thách trong grid ngữ nghĩa.....	12
3.4	Kiến trúc grid ngữ nghĩa	14
II.	KIẾN TRÚC S-OGSA CỦA LƯỚI NGỮ NGHĨA	14
1.	Giới thiệu chung.....	14
2.	Các kiến trúc lưới tham khảo	15
2.1	Các nguyên lý thiết kế kiến trúc Lưới ngữ nghĩa tham khảo	16
3.	Mô hình và khả năng của S-OGSA.....	17
3.1	Mô hình S-OGSA	17
3.2	Khả năng của S-OGSA	18
3.2.1	Các dịch vụ cung cấp ngữ nghĩa	18
3.2.2	Các dịch vụ Lưới Ngữ Nghĩa.....	20

3.3	Ví dụ về khả năng và mô hình S-OGSA trong ngữ cảnh hệ thống điều khiển truy xuất có ngữ nghĩa	20
4.	Cơ chế của S-OGSA	21
4.1	Đối xử các thực thể tri thức và việc kết hợp ngữ nghĩa như là những tài nguyên của lưới	21
4.2	Các dịch vụ S-Stateful	22
III.	ỨNG DỤNG GRID NGỮ NGHĨA MYEXPERIMENT	23
IV.	PHỤ LỤC PHÂN CÔNG CÔNG VIỆC	31

I. GRID NGỮ NGHĨA

1. Giới thiệu

Việc phát triển và nghiên cứu khoa học luôn luôn gồm một số lượng lớn con người ở nhiều cấp độ khác nhau, làm việc ở những vị trí khác nhau cả độc lập hay cộng tác và sử dụng một lượng lớn các kiến thức. Tuy nhiên trong những năm gần đây, có một số sự thay đổi quan trọng trong bản chất cũng như tiến trình nghiên cứu. Đặc biệt, nhấn mạnh trong việc cộng tác giữa các nhóm lớn, sử dụng các kỹ thuật xử lý thông tin cấp cao và nhu cầu chia sẻ kết quả nghiên cứu và quan sát giữa các thành viên nằm ở các khoảng cách vật lý cách xa nhau.

Những xu hướng này có nghĩa là các nhà nghiên cứu ngày càng trông cậy nhiều hơn vào máy tính và công nghệ thông tin như là một phần không thể thiếu trong hoạt động nghiên cứu hàng ngày. Hiện tại thì kỹ thuật thông tin chủ yếu là email và web.

Hệ thống Grid ra đời nhằm hỗ trợ việc chia sẻ tài nguyên một cách an toàn, linh hoạt và có tổ chức. Cộng đồng Grid đã phát triển một kiến trúc chung cho các hệ thống Grid, gọi là OGSA (Kiến trúc Grid hướng dịch vụ mở). Kiến trúc này đã giải quyết nhu cầu chuẩn hóa Grid bằng cách định nghĩa một tập các chức năng và hành vi của một hệ thống Grid.

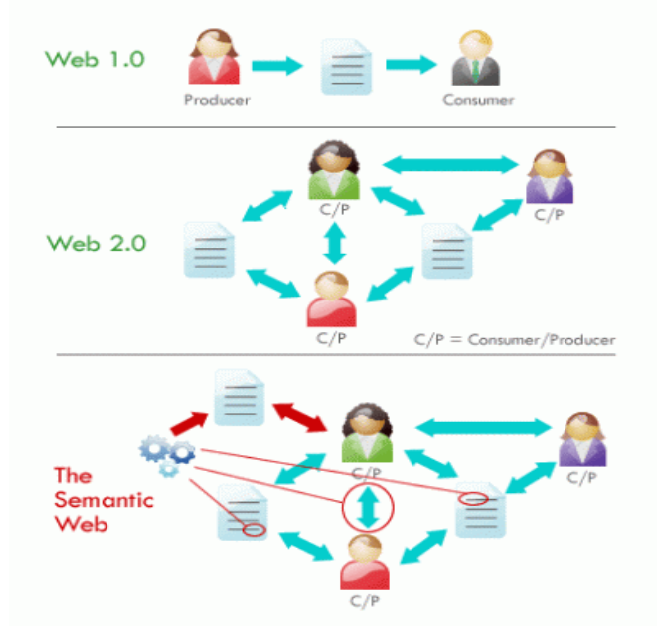
Tuy nhiên, trong các dịch vụ của Grid, tri thức biểu diễn bởi các metadata được xử lý một cách rất phức tạp, không tường minh, nằm ẩn sâu trong thư viện mã của hệ thống. Điều này dẫn đến việc khó chia sẻ tri thức cũng như việc sử dụng hiệu quả tài nguyên giữa các hệ thống Grid với nhau. Vì vậy phát sinh yêu cầu là phát triển các chuẩn công nghệ chung để biểu diễn thông tin và cho phép máy tính có thể hiểu được một số thông tin trên Web, hỗ trợ tìm kiếm thông minh hơn, hỗ trợ việc khám phá, tách chiết thông tin, tích hợp dữ liệu và tự động hóa một số công việc thay cho con người.

Hệ thống Grid ngữ nghĩa ra đời nhằm cung cấp thông tin giàu ngữ nghĩa cho các tài nguyên của Grid, giúp xây dựng các dịch vụ Grid thông minh hơn và giúp việc chia sẻ tài nguyên giữa các hệ thống Grid trở nên dễ dàng.

Grid ngữ nghĩa là sự mở rộng của Grid hiện tại trong đó thông tin và dịch vụ được cung cấp nghĩa được định nghĩa tốt (well-defined meaning) thông qua các mô tả mà máy tính có thể xử lý được, cho phép chúng được sử dụng bởi con người và máy để con người và máy làm việc cộng tác với nhau.

Grid ngữ nghĩa chủ yếu dựa vào công nghệ Web ngữ nghĩa trước đó, đồng thời nó tuân thủ theo mô hình S-OGSA (mở rộng từ OGSA) để thiết kế các thành phần và ứng dụng ngữ nghĩa.

2. Web ngữ nghĩa



Hình 1. Web ngữ nghĩa

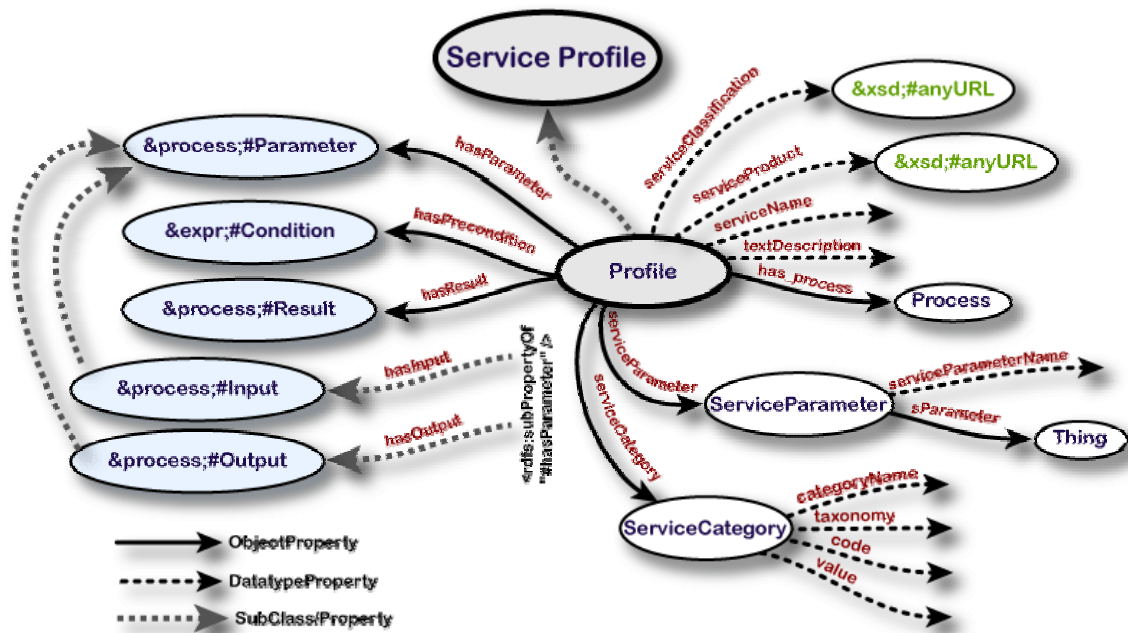
World Wide Web, gọi tắt là Web hay WWW, được phát minh và đưa vào sử dụng vào khoảng năm 1990, 1991 bởi viện sĩ Viện Hàn lâm Anh Tim Berners-Lee. Từ đó đến nay nó đã phát triển một cách rộng khắp và trở thành một dịch vụ không thể thiếu trên Internet. Ban đầu nó đơn thuần chỉ là dịch vụ chia sẻ thông tin nhưng ngày nay nó dần dần tiến tới một kỹ nguyên mới với khái niệm Web 2.0, cho phép người dùng có thể sửa đổi thông tin trực tiếp mà không phải thông qua quyền quản trị, giúp cho con người trên mọi vùng lãnh thổ tiến lại gần nhau hơn.

Tuy nhiên vấn đề của Web hiện tại là thông tin được biểu diễn dưới dạng văn bản thô mà chỉ con người mới có thể đọc hiểu được. Điều này thúc đẩy sự ra đời của ý tưởng Web có ngữ nghĩa (Semantic Web), một thế hệ mới của Web, mà theo Tim Berners-Lee nó là sự mở rộng của Web hiện tại mà trong đó thông tin được định nghĩa rõ ràng sao cho con người và máy tính có thể cùng làm việc với nhau một cách hiệu quả hơn.

Xét về mặt bản chất, Semantic Web chỉ là một công cụ để con người cũng như máy tính sử dụng để biểu diễn thông tin, hay nói chính xác hơn thì Semantic Web chỉ là một dạng dữ liệu trên Web. Khác với các dạng thức dữ liệu được trình bày trong HTML, dữ liệu trong Semantic Web được đánh dấu, phân lớp, mô hình hóa, được bổ sung thêm các thuộc tính, các mối liên hệ... theo các lĩnh vực cụ thể, qua đó giúp cho các phần mềm máy tính có thể hiểu được dữ liệu và tự động xử lý được những dữ liệu đó. Chẳng hạn, với cụm từ “phần mềm này có thể chạy trên hệ điều hành Unix”, trong không gian Semantic Web, nó cũng có thể được suy luận tới: “phần mềm này cũng có thể chạy trên các hệ điều hành khác tương tự như Unix như Linux, Ubuntu, SunOS, ...”

Mục tiêu của Web có ngữ nghĩa là phát triển các chuẩn chung và công nghệ cho phép máy tính hiểu được nhiều hơn thông tin trên Web nhằm hỗ trợ tốt hơn trong việc khám phá thông tin, tích hợp dữ liệu và tự động hóa các công việc. Phần tiếp theo sẽ trình bày các khái niệm và công nghệ liên quan đến Web có ngữ nghĩa.

2.1 Ontology



Hình 2. Một ví dụ về ontology

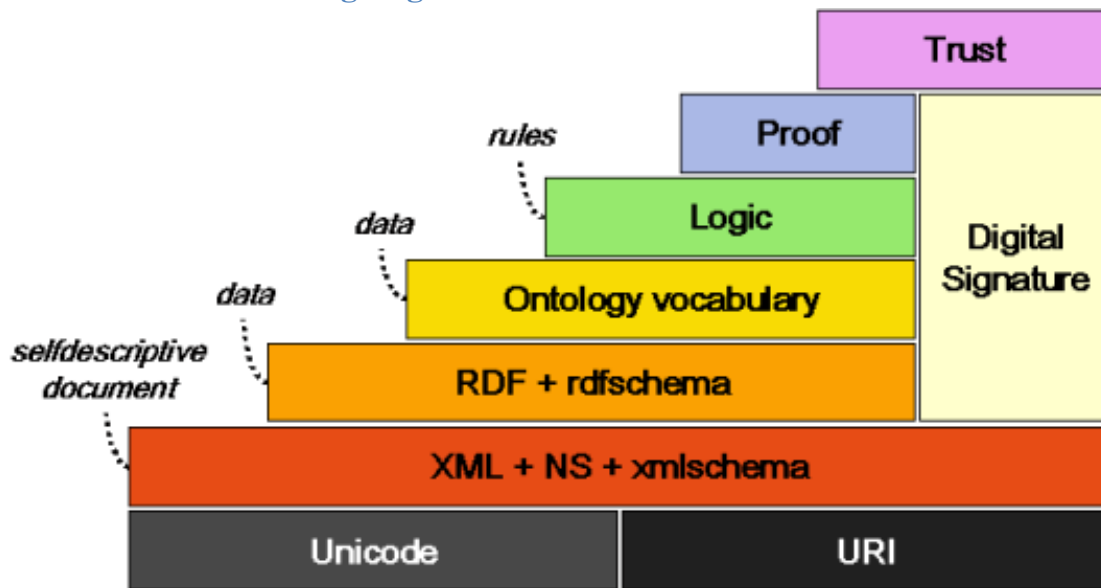
Trong ngữ cảnh của Web có ngữ nghĩa, Ontology đóng vai trò then chốt trong việc cung cấp ngữ nghĩa mà máy có thể hiểu được cho các tài nguyên của Web ngữ nghĩa.

Ontology là một thuật ngữ mượn từ triết học nhằm chỉ khoa học mô tả các loại thực thể trong thế giới thực và cách chúng liên kết với nhau. Trong khoa học máy tính, một cách khái quát, Ontology là "một biểu diễn của sự khái niệm hoá chung được chia sẻ" của một miền nhất định.

Nó cung cấp một bộ từ vựng chung bao gồm các khái niệm, các thuộc tính quan trọng và các định nghĩa về các khái niệm và các thuộc tính này. Ngoài bộ từ vựng, Ontology còn cung cấp các ràng buộc, đôi khi các ràng buộc này được coi như các giả định cơ sở về ý nghĩa mong muốn của bộ từ vựng, nó được sử dụng trong một miền mà có thể được giao tiếp giữa người và các hệ thống ứng dụng khác.

Ontology (được hiểu là bộ từ điển) có thể được hiện thực bằng XML, XML Namespace, XML Schema, RDF, RDF Schema và OWL.

2.2 Kiến trúc của web ngữ nghĩa



Hình 3. Kiến trúc của web ngữ nghĩa

2.2.1 Unicode

Unicode là bảng mã chuẩn chung có đủ các ký tự để thống nhất sự giao tiếp trên tất cả các quốc gia, nhằm đáp ứng tính nhất quán toàn cầu của web.

2.2.2 URI

URI (Uniform Resource Identifier) là kí hiệu nhận dạng Web đơn giản. Cụ thể, nó là một chuỗi cho phép nhận dạng tài nguyên Web như các chuỗi bắt đầu với "http:" hoặc "ftp:" mà chúng ta thường thấy trên World Wide Web.

2.2.3 XML

XML (Extensible Markup Language) là ngôn ngữ đánh dấu mở rộng được phát triển dựa trên tính đơn giản, dễ sử dụng của ngôn ngữ HTML nhưng cho phép định nghĩa các thẻ (tag) theo nhu cầu sử dụng để mô tả các tài liệu có cấu trúc mà ngôn ngữ HTML không làm được. XML được thiết kế để cho phép máy tính có thể trao đổi tài liệu với nhau thông qua Web mà không làm mất đi ý nghĩa của dữ liệu.

Các thực thể (entity) markup chính trong XML là các đơn vị (element). Chúng thông thường bao gồm một tag mở và một tag đóng. Ví dụ <person> và </person>.

Các element có chứa các element khác hay text. Các element nên được xếp lồng nhau, tag mở và tag đóng của element con phải nằm trong tag mở và tag đóng của element cha.

Mỗi XML document phải có chính xác một root element.

Các element có thể chứa thuộc tính với giá trị nào đó, có định dạng là: "từ = giá trị" bên trong tag của một element. Ví dụ <person name="John">.

Dưới đây là một ví dụ về XML:

```
<?xml version="1.0"?>
<Students>
    <person name="Võ Bảo Hùng">
        <phone>123456</phone>
    </Students>
```

2.2.4 XML Schema

XML Schema là một ngôn ngữ được dùng để định nghĩa cấu trúc của một tài liệu XML như là phần tử nào xuất hiện trong tài liệu, quan hệ cha con giữa các phần tử, kiểu dữ liệu của các phần tử,... XML Schema sử dụng cú pháp của ngôn ngữ XML và được xem như một sự thay thế cho DTD (Document Type Definition) đã lỗi thời và khó sử dụng.

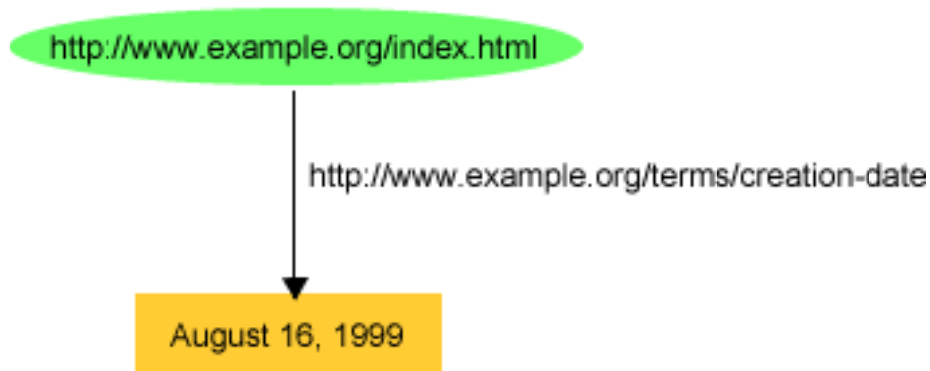
XML Schema mở đầu bằng khai báo theo chuẩn XML, tiếp theo dùng tiếp đầu ngữ xsd: để khai báo không gian tên XML Schema, theo cú pháp sau:

```
<?xml version = "1.0" ?>
<xsd:schema xmlns:xsd = http : // www.w3.org/1999/XMLSchema>
    .....
</xsd:schema>
```

2.2.5 RDF

Có thể nói rằng RDF (Resource Description Framework) chính là nền tảng và là linh hồn của Web có ngữ nghĩa. RDF là một ngôn ngữ được dùng để mô tả thông tin về những tài nguyên trên Web và mô tả ngữ nghĩa của những thông tin ấy theo cách mà máy có thể hiểu được. RDF thích hợp trong những ứng dụng mà ở đó thông tin cần được xử lý bởi máy tính chứ không phải con người. RDF cung cấp một framework chung cho việc biểu diễn thông tin này vì thế nó có thể được trao đổi giữa các ứng dụng mà không làm mất đi ý nghĩa của thông tin. RDF mô tả tài nguyên trên Web thông qua URI (Uniform Resource Identifier).

Trong RDF, thông tin được thể hiện bởi bộ ba subject - predicate – object hay (Subject, Predicate, Object). Ví dụ phát biểu “<http://www.example.org/index.html> has a creation-date whose value is August 16, 1999” sẽ được biểu diễn dưới dạng bộ ba như sau (“<http://www.example.org/index.html>”, “creation-date”, “August 16, 1999”). Các bộ ba này có thể được biểu diễn dưới dạng đồ thị, gọi là đồ thị RDF (RDF Graph). Tất cả các phần tử trong bộ ba là các tài nguyên được xác định duy nhất bởi các URI, riêng thành phần object, nó có thể là URI, là hằng chuỗi hoặc là một con số.



Hình 4. Mô tả việc biểu diễn thông tin đơn giản bằng bộ ba Subject-Predicate-Object

Cú pháp chính qui cho RDF là RDF/XML. Nó là sự kết hợp giữa cú pháp của ngôn ngữ XML và khả năng mô tả tài nguyên thông qua các URI của RDF. Giữa RDF/XML và đồ thị RDF có sự tương đương 1:1. Ví dụ, để biểu diễn đồ thị RDF ở hình 1, ta dùng cú pháp của RDF/XML như sau:

```
<?xml version = "1.0" ?>
<rdf:RDF      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:extms=http://www.example.org/terms/      >
  <rdf:Description rdf:about= http://www.example.org/index.html>
    <extms:creation-date> August 16, 1999</ extms:creation-date >
  </rdf:Description>
</rdf:RDF>
```

2.2.6 RDF Schema

RDF Schema (RDFS) là sự mở rộng của RDF để cho phép mô tả sự phân loại của các lớp (classes) và các thuộc tính (properties). RDF Schema định nghĩa các classes và properties để mô tả các classes, properties và các tài nguyên khác. Nó cũng có thể xem như là một sự mở rộng ngữ nghĩa của RDF để cung cấp những cơ chế cho phép mô tả các nhóm của các tài nguyên liên quan và mối quan hệ giữa các tài nguyên này.

Trong RDFS, classes là một nhóm các tài nguyên có liên quan với nhau. Bảng 1 dưới đây liệt kê danh sách các classes trong RDFS

Element	Class of	rdfs:subClassOf	rdf:type
rdfs:Resource	all resources	rdfs:Resource	rdfs:Class
rdfs:Class	all classes	rdfs:Resource	rdfs:Class
rdfs:Literal	literal values	rdfs:Resource	rdfs:Class

rdfs:Datatype	datatypes	rdfs:Class	rdfs:Class
rdf:XMLLiteral	XML literal values	rdfs:Literal	rdfs:Datatype
rdf:Property	properties	rdfs:Resource	rdfs:Class
rdf:Statement	statements	rdfs:Resource	rdfs:Class
rdf:List	lists	rdfs:Resource	rdfs:Class
rdfs:Container	containers	rdfs:Resource	rdfs:Class
rdf:Bag	unordered containers	rdfs:Container	rdfs:Class
rdf:Seq	ordered containers	rdfs:Container	rdfs:Class
rdf:Alt	containers of alternatives	rdfs:Container	rdfs:Class
rdfs:Container MembershipProperty	rdf:_1... Properties expressing membership	rdf:Property	rdfs:Class

Bảng 1. Danh sách các class trong RDF

Properties trong RDFS chính là quan hệ giữa các Subjects và Objects trong RDF. Danh sách các properties được liệt trong trong bảng 2.

Element	Relates	rdfs:domain	rdfs:range
rdfs:range	restricts subjects	rdf:Property	rdfs:Class
rdfs:domain	restricts objects	rdf:Property	rdfs:Class
rdf:type	instance of	rdfs:Resource	rdfs:Class
rdfs:subClassOf	subclass of	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	subproperty of	rdf:Property	rdf:Property
rdfs:label	human readable label	rdfs:Resource	rdfs:Literal
rdfs:comment	human readable comment	rdfs:Resource	rdfs:Literal
rdfs:member	container membership	rdfs:Resource	rdfs:Resource

rdf:first	First element	rdf:List	rdfs:Resource
rdf:rest	rest of list	rdf:List	rdf:List
rdf:_1, rdf:_2, ...	container membership	rdfs:Container	rdfs:Resource
rdfs:seeAlso	further information	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	definition	rdfs:Resource	rdfs:Resource
rdf:value	For structured values	rdfs:Resource	rdfs:Resource
rdf:object	object of statement	rdf:Statement	rdfs:Resource
rdf:predicate	predicate of statement	rdf:Statement	rdfs:Resource
rdf:subject	subject of of statement	rdf:Statement	rdfs:Resource

Bảng 2. RDFs Properties

2.2.7 OWL

OWL (Web Ontology Language) là một sự mở rộng từ RDF và RDFS. Mục đích chính của OWL là đưa khả năng suy luận vào Web có ngữ nghĩa. OWL có ba loại: OWL Lite, OWL DL và OWL Full. Mỗi loại OWL sẽ có những đặc tính riêng và do đó sẽ phù hợp trong ngữ cảnh của một ứng dụng nào đó.

2.2.8 Logic

Việc biểu diễn tài nguyên dưới dạng các bộ từ vựng ontology có mục đích là để máy có thể lập luận được. Mà cơ sở lập luận chủ yếu dựa vào logic. Chính vì vậy mà các ontology được ánh xạ sang logic, cụ thể là logic mô tả để có thể hỗ trợ lập luận

2.2.9 Proof

Tầng này đưa ra các luật để suy luận các thông tin mới dựa trên những thông tin đã có. Ví dụ: A là cha của B, A là em trai C thì khi đó ta có thông tin mới C là bác của B. Để có được các suy luận này thì cơ sở là FOL (First-Order-Logic). Và tầng này hiện nay các nhà nghiên cứu đang xây dựng các ngôn ngữ luật như SWRL, RuleML

2.2.10 Trust

Đảm bảo tính tin cậy của các ứng dụng trên Web ngữ nghĩa

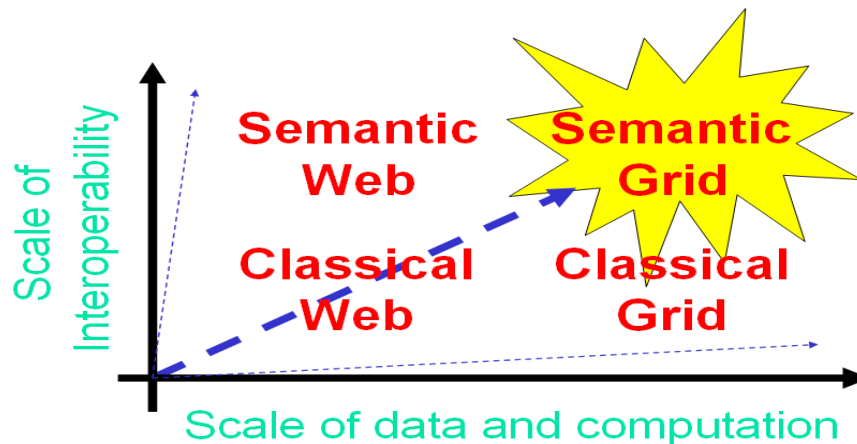
3 Grid ngữ nghĩa

3.1 Khái niệm

Grid ngữ nghĩa là một hệ thống mở trong đó người dùng, các thành phần phần mềm và các tài nguyên tính toán (được sở hữu bởi nhiều stakeholder khác nhau) tham gia và tiến hành

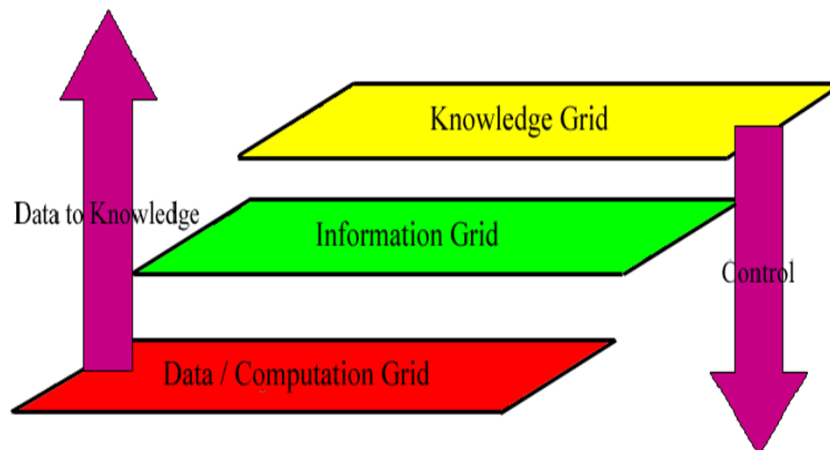
trên nền tảng cơ bản. Có thể có một cấp độ tự động cao hơn mà hỗ trợ việc cộng tác và tính toán linh động trong môi trường toàn cầu. Hơn nữa, môi trường này nên được cá nhân hóa cho từng thành viên riêng biệt và tạo ra tương tác liền mạch giữa các thành phần phần mềm cũng như những người sử dụng có liên quan khác.

Grid ngữ nghĩa là sự mở rộng của Grid hiện tại trong đó thông tin và dịch vụ được cung cấp nghĩa được định nghĩa tốt (well-defined meaning) thông qua các mô tả mà máy tính có thể xử lý được, cho phép chúng được sử dụng bởi con người và máy để con người và máy làm việc cộng tác với nhau.



Hình 5. Ý tưởng grid ngữ nghĩa

3.2 Ba lớp khái niệm trong hạ tầng tính toán



Hình 6. Ba lớp khái niệm trong hạ tầng tính toán

- Dữ liệu / Tính toán (data/computation):

Lớp này xử lý cách mà các tài nguyên tính toán được định vị, lập lịch và thực thi; cũng như cách mà dữ liệu được di chuyển giữa các tài nguyên đang xử lý khác nhau. Nó được thiết lập

để mà có khả năng xử lý với 1 lượng dữ liệu lớn, cung cấp đường mạng nhanh và hiển thị các tài nguyên khác nhau như là một siêu máy tính đơn.

Lớp này có thể được xây dựng trên lớp ‘grid fabric’.

- Thông tin (information):

Lớp này xử lý cách mà thông tin được hiển thị, lưu trữ, truy xuất, chia sẻ và bảo trì. Ở đây, thông tin được hiểu như là dữ liệu có nghĩa. Ví dụ, đặc điểm của 1 số nguyên khi biểu diễn nhiệt độ của tiến trình phản ứng, nhận diện 1 chuỗi như là tên của 1 ai đó.

- Kiến thức (knowledge):

Lớp này liên quan tới cách mà kiến thức được thu được, sử dụng, truy hồi, xuất bản và bảo trì để hỗ trợ e-Scientist đạt được các mục tiêu riêng biệt. Ở đây, kiến thức được hiểu như là thông tin được ứng dụng để đạt được mục đích, giải quyết vấn đề hoặc đưa ra quyết định. Ví dụ, nhận diện 1 tác vụ trong nhà máy mà trong ngữ cảnh hiện tại nhiệt độ phản ứng yêu cầu phải dừng tiến trình.

3.3 Các yêu cầu và thử thách trong grid ngữ nghĩa

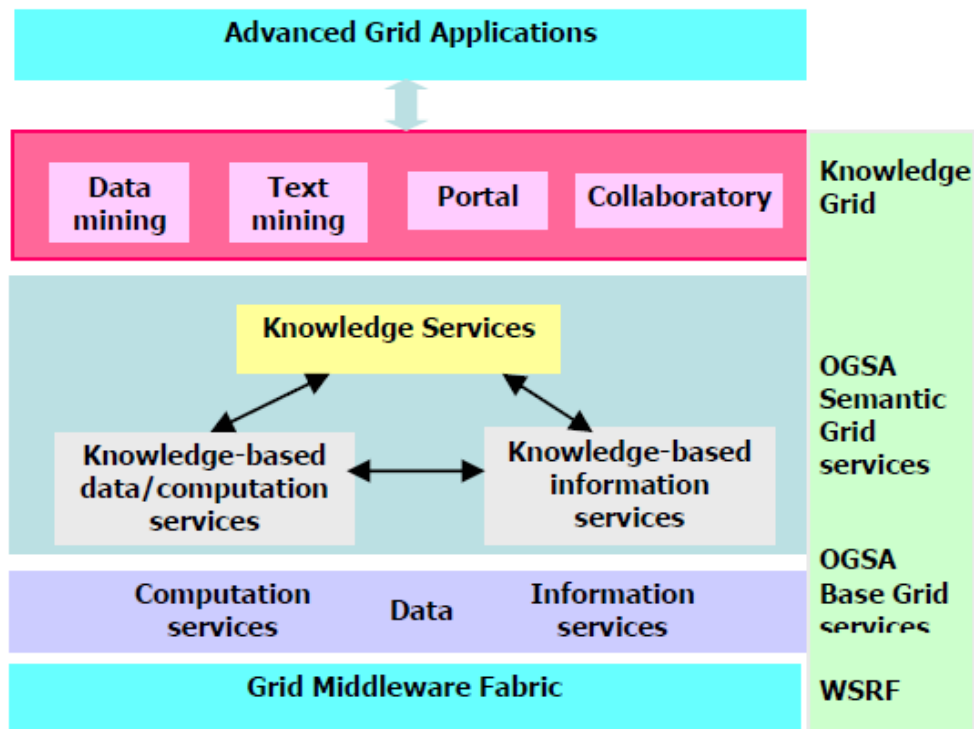
- *Sử dụng, khám phá và đặc tả các tài nguyên*: Hệ thống phải lưu trữ và xử lý 1 lượng lớn các thông tin phân bố theo 1 các hiệu quả và tiết kiệm thời gian, có thể thông qua sự kế hợp các tài nguyên, và vì vậy có thể xác định thông tin, dịch vụ, các tài nguyên tính toán và các công cụ khả dụng trong Grid, ... Hệ thống cũng phải phát hiện và định vị các tài nguyên này 1 cách hiệu quả và thương lượng để truy cập. Nó cũng cần yêu cầu sự tổng quát và xử lý của các đặc tả công việc và các lệnh và lập kế hoạch động cho việc sử dụng các tài nguyên để đáp ứng được chất lượng của các yêu cầu dịch vụ và đạt được tính khả dụng của các tài nguyên.
- *Đặc tả và ban hành các tiến trình*: Để hỗ trợ việc tạo ra có tổ chức ảo cung cấp dịch vụ, hệ thống cần các đặc tả (giống như là những workflow) để khởi tạo việc kết hợp của nhiều tài nguyên và các cơ chế để tạo và điều hành những tài nguyên theo cách thức phân bố.
- *Hành xử độc lập*: Hệ thống nên tự cấu hình để đáp ứng các nhu cầu đa người dùng khi hoàn cảnh thay đổi động, và tự sửa chữa khi có những lỗi xảy ra (hệ thống dường như là đáng tin cậy nhưng thực chất nó có thể che giấu những lỗi và các thụ lý ngoại lệ ở nhiều cấp khác nhau). Hệ thống cũng nên hỗ trợ việc phát triển thêm các thông tin và dịch vụ mới.
- *Bảo mật và đáng tin*: Các cơ chế xác thực, mã hóa và những yêu cầu riêng tư khác với sự tham gia của đa tổ chức và 1 yêu cầu tổng quát cho những thứ này là chúng nên được xử lý với việc can thiệp thủ công tối thiểu. Liên quan đến vấn đề này là những vấn đề về trao đổi: những người tham gia khác nhau yêu cầu phải được giữ quyền sở hữu những thông tin và năng lực xử lý của họ, cho phép những người khác truy cập vào dưới những điều kiện và thỏa thuận hợp lý.

- *Chú thích:* Từ việc log lại mẫu để công bố việc phân tích, việc có những chú thích để làm rõ thêm những đặc tả của các thông tin số là cần thiết. Thông tin dạng meta này có thể áp dụng cho dữ liệu, thông tin hay là kiến thức, phụ thuộc vào cách diễn giải đã được thống nhất trước. Lý tưởng là trong nhiều trường hợp, những chú thích này có thể được rút trích ra 1 cách tự động. Chúng cũng hỗ trợ những ký hiệu quan trọng về nguồn gốc, nơi mà những thông tin cần thiết được lưu trữ để mà có thể lặp lại các thực nghiệm, sử dụng lại các kết quả và cung cấp những bằng chứng rằng thông tin này thật sự được tạo ra tại thời điểm này (sau này có thể có 1 bên thứ 3 tham gia vào). Tuy nhiên thì việc chứa những chú thích chỉ là phân nửa công việc. Chúng ta cũng cần sử dụng chúng. Ví dụ như: tìm kiếm các bài báo, tìm người hay tìm lại một thiết kế thực nghiệm trước (những truy vấn này có thể bao gồm suy luận được rút ra), chú thích cho những phân tích đã được đưa lên và cấu hình 1 không gian thông minh để tương thích với người sử hữu. Những chú thích này cũng có thể được phân tán hay cộng tác.
- *Tích hợp thông tin:* Khả năng tạo ra những câu truy vấn có nghĩa trong không gian lưu trữ thông tin riêng rẽ và sử dụng thông tin theo cách có thể hoặc không thể dự đoán trước được, yêu cầu việc kết hợp các thông tin. Ví dụ, có thể bao gồm việc ánh xạ các thuật ngữ được sử dụng trong các lĩnh vực khác nhau. Điều này là kỹ thuật cơ bản trong công nghệ Semantic Web.
- *Đồng bộ các dòng thông tin và hợp nhất chúng lại:* Bên cạnh việc lưu trữ ổn định, có thêm 1 yêu cầu khi làm việc với các dòng dữ liệu theo thời gian. Những dòng dữ liệu này có thể là dữ liệu từ các thiết bị, video hay bất cứ dòng dữ liệu nào khác như là kết quả từ sự tương tác. Cả dòng dữ liệu sống và tái diễn đều có 1 vai trò nhất định, đặc biệt ở những nơi mà dòng dữ liệu này được làm phong phú bởi các dữ liệu meta theo thời gian tương ứng. Những thông báo cũng là dòng dữ liệu, việc có thêm 1 thông tin mới sẽ phát sinh thông báo đến người sử dụng và khởi tạo 1 quá trình xử lý tự động. Việc hỗ trợ là cần thiết cho cả việc thu lại và tái diễn lại các dòng dữ liệu và làm cho việc tổng hợp các thông tin mới từ chúng. Hơn nữa thông tin cần để có thể kết hợp từ nhiều nguồn khác nhau theo 1 cách không thể dự đoán trước được tùy theo nhu cầu của người sử dụng. Những đặc tả nguồn gốc và thông tin, những chú thích, sẽ được kết hợp theo 1 cách có nghĩa
- *Hỗ trợ ra quyết định tùy theo ngữ cảnh:* Thông tin cần để được biểu diễn cho người sử dụng vào đúng thời điểm, đúng định dạng trên đúng thiết bị và với cấp độ xâm phạm vừa phải. Đó là nhiệm vụ của hệ thống thông tin tương thích và được nhấn mạnh bởi sự mềm dẻo của năng lực của Grid.
- *Cộng đồng:* Những người sử dụng có thể định dạng, bảo trì và giải tán các nhóm thí nghiệm với các điều kiện thành viên và các vai trò thực thi ngặt nghèo. Điều này bao gồm việc xác định tập các cá nhân trong 1 tổ chức ảo thông qua các công cụ cộng tác và khai thác thông tin về nhóm thông qua các nguyên tắc.
- *Môi trường thông minh:* Môi trường nên biểu thị một cấp độ thông minh ở xung quanh. Ví dụ, công cụ phát hiện mẫu (mã vạch hoặc là RFID), các nhà khoa học có thể sử dụng

các thiết bị di động để đánh dấu, nút Access Grid có thể làm được các việc như nhận dạng giọng nói, việc ảnh hóa có thể sử dụng các thiết bị hiển thị khác nhau.

- *Thiết kế hình và triển khai:* Các ứng dụng lưới nên có thể chạy được bởi những người thường, không phải là chuyên gia.
- *Tích hợp với hệ thống IT đã có:* Thách thức về mặt tương thích của việc cùng làm với những xử lý nghiệp vụ với hệ thống quản lý việc học hỏi cần được định vị theo 1 cách thức thỏa mãn hơn vì môi trường semantic grid là lĩnh vực nhàm chán.

3.4 Kiến trúc grid ngữ nghĩa



Hình 7. Kiến trúc grid ngữ nghĩa

Kiến trúc của grid ngữ nghĩa chủ yếu dựa vào kiến trúc S-OGSA là mở rộng của kiến trúc OGSA.

II. KIẾN TRÚC S-OGSA CỦA LƯỚI NGỮ NGHĨA

1. Giới thiệu chung

Mục đích của Lưới là hỗ trợ tính bảo mật, an toàn, linh động và sự chia sẻ tài nguyên giữa các nhà cung cấp dịch vụ cho việc tính toán phân tán. Do đó, cơ sở hạ tầng của Lưới là phải cho phép thu thập bất kỳ loại tài nguyên nào như là: việc tính toán, nơi lưu trữ, các tập dữ liệu, các

thư viện số, ... để dễ dàng tạo nên những tổ chức ảo (Virtual Organization - VOs) để có thể làm việc với nhau để giải quyết các vấn đề.

Tính toán lưới là một dạng phổ biến nhất của Lưới, có khả năng tính toán lớn để hỗ trợ việc tính toán cho các ứng dụng phân tích. Trong đó các Lưới này được ảo hóa thành các máy tính đơn ảo đối với ứng dụng, các Lưới dữ liệu thể hiện một kho dữ liệu đơn ảo mà thực ra nó có thể được phân bố ở khắp nơi. Các cổng thông tin cung cấp cách thức cho các nhà phát triển ứng dụng và người dùng để gửi công việc cần tính toán hay các truy vấn của họ. Để dễ dàng phát triển các tất cả các ứng dụng của Lưới, cộng đồng Lưới, thông qua Global Grid Forum, đã phát triển một kiến trúc tham khảo đó là kiến trúc dịch vụ lưới mở (Open Grid Service Architecture) viết tắt là OGSA. Kiến trúc này đã chỉ ra sự cần thiết cho việc chuẩn hóa bằng cách định nghĩa một tập các ứng xử và khả năng cơ bản được xem là những mối quan tâm quan trọng trong hệ thống Lưới.

Lưới phụ thuộc vào sự hiểu biết các tài nguyên có sẵn và khả năng của chúng, làm cách nào để tìm ra chúng và kết hợp chúng lại với nhau một cách tốt nhất. Do đó, các phần mềm Lưới trung gian, các ứng dụng Lưới được chúng hỗ trợ, đã phát triển các siêu dữ liệu để mô tả các tài nguyên trong tất cả các cấu tạo, các tổ chức ảo (Vos), các chính sách của chúng để chỉ dẫn cho chúng... và kết hợp với các tri thức để áp dụng các siêu dữ liệu đó một cách thông minh.

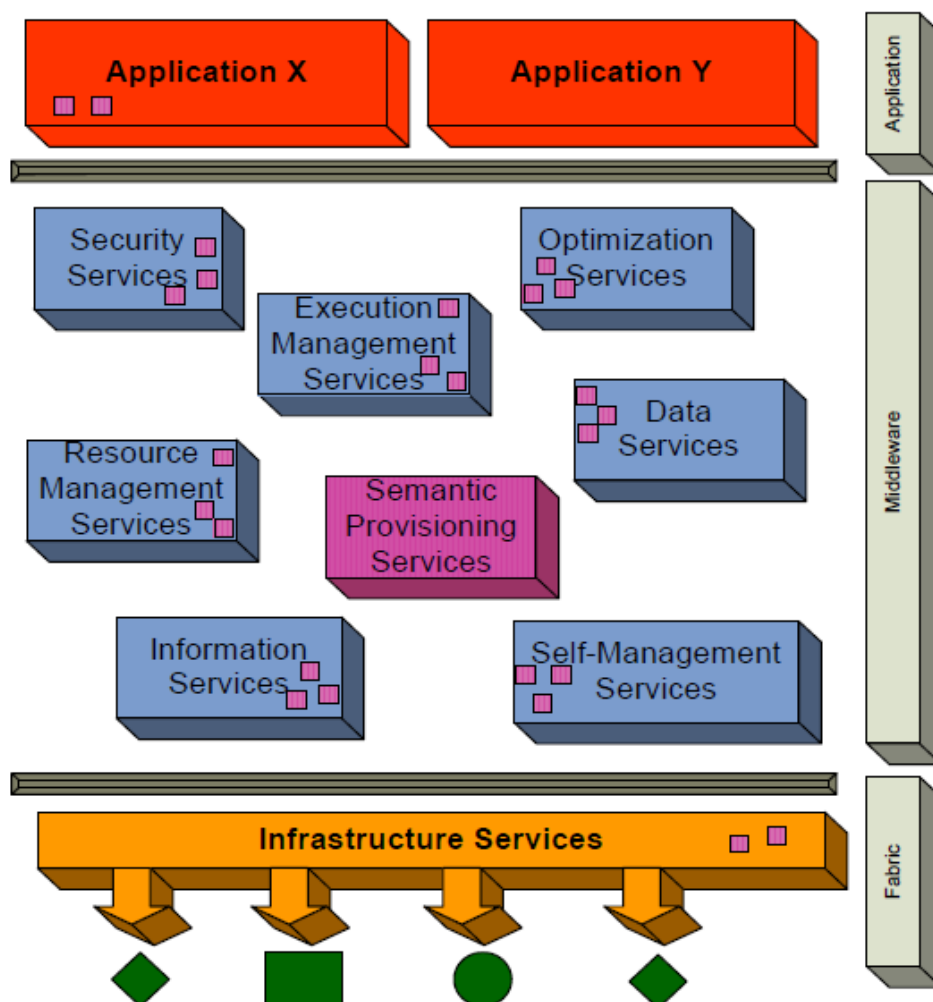
2. Các kiến trúc lưới tham khảo

Các nhà nghiên cứu Lưới đã cố gắng cung cấp các định nghĩa chính xác cho Lưới. Sau đây là các đặc tính của Lưới:

- Lưới tập trung vào việc chia sẻ các tài nguyên phân bố theo cách được điều khiển tốt và công bằng trong việc tạo các virtual pool. Người sử dụng của pool có thể biết rất ít hoặc không biết gì về trạng thái thực, kiểu và tính năng của các tài nguyên.
- Để đạt được (a), giả sử rằng hai thực thể chủ yếu trong môi trường là Người sử dụng và Tài nguyên, ở mức virtual pool là cần thiết. Việc ánh xạ các khái niệm trừu tượng này của Người sử dụng (ví dụ: các định danh toàn cục được thể hiện bằng chứng chỉ) tới các thực thể cục bộ (ví dụ: các luật trong cơ sở dữ liệu) và Tài nguyên (ví dụ: một hạng mục dữ liệu ảo) tới các tài nguyên thực (ví dụ: các tập tin nằm trong ổ đĩa) tại thời gian của việc tối ưu hóa tài nguyên là cần thiết.

OGSA là kết quả của nỗ lực chuẩn hóa, mà bây giờ đã được chấp nhận bởi hội đồng chuẩn của Lưới có tên là Global Grid Forum (GGF). OGSA nhắm tới định nghĩa một tập cơ bản các ứng xử và khả năng cho hệ thống Lưới.

OGSA đưa ra cách nhìn logic kiến trúc ba tầng của môi trường phân tán bằng việc sử dụng các hệ thống Lưới, như hình 1 dưới đây:



Hình 8 Kiến trúc 3 tầng của Lưới Ngữ Nghĩa

Tầng cơ bản là *Fabric* trong đó các kiểu tài nguyên khác nhau được ảo hóa được ảo hóa thông qua các dịch vụ web (web-services). OGSA bao gồm các tầng giữa, để tạo thành các danh mục dịch vụ được gọi là *Grid middleware*. Tầng trên cùng là các ứng dụng (application). Các ứng dụng sẽ sử dụng các Grid middle để thực hiện các hoạt động của chúng.

2.1 Các nguyên lý thiết kế kiến trúc Lưới ngữ nghĩa tham khảo

Lưới ngữ nghĩa hiện tại thiếu một kiến trúc tham khảo hay một nền tảng ngữ nghĩa cho việc thiết kế các thành phần hay các ứng dụng cho Lưới ngữ nghĩa. Vì vậy kiến trúc trúc của Lưới Ngữ Nghĩa (S-OGSA) đã được phát triển dựa trên kiến trúc OGSA theo sáu nguyên lý sau đây:

1. **Tính tinh giản của các thành phần kiến trúc:** kiến trúc Lưới mới cần phải đủ độ tinh giản cần thiết để giảm thiểu những ảnh hưởng lên kiến trúc Lưới và các công cụ truyền thống.

2. **Tính có thể mở rộng của framework:** định nghĩa một kiến trúc có thể mở rộng và tùy biến thay vì định nghĩa một kiến trúc đầy đủ và tổng quát.
3. **Tính thống nhất của cơ chế:** Lưới Ngữ nghĩa cũng là Lưới , vì thế tất cả các thực thể của Lưới Ngữ Nghĩa cũng đều có tính chất giống như các thực thể của Lưới thông thường, các tính chất đó là:
 - a. Các tài nguyên của Lưới Ngữ nghĩa có thể quản lý được. Các mô tả ngữ nghĩa có trạng thái và thời gian sống.
 - b. Lưới Ngữ Nghĩa phải hỗ trợ các dịch vụ có trạng thái (stateful) và không trạng thái (stateless) giống Lưới thông thường.
 - c. Các dịch vụ tri thức (Knowledge services) trong Lưới Ngữ Nghĩa cũng tuân thủ các đặc tả của các dịch vụ Lưới thông thường.
4. **Tính đa dạng của các khả năng ngữ nghĩa:** trong hệ thống Lưới Ngữ Nghĩa các thực thể có thể có nhiều mức khả năng hỗ trợ ngữ nghĩa khác nhau:
 - a. Các thực thể không hỗ trợ ngữ nghĩa.
 - b. Các thực thể hiểu ngữ nghĩa nhưng không thể xử lý ngữ nghĩa.
 - c. Các thực thể hiểu ngữ nghĩa và có khả năng xử lý ngữ nghĩa đầy đủ hoặc bán đầy đủ.
5. **Tính đa dạng của việc biểu diễn ngữ nghĩa:** ngữ nghĩa của các thực thể có thể được mô tả , biểu diễn ở nhiều dạng khác nhau như : text , logic, ontology ,rule.
6. **Tính dễ chuyển đổi của các dịch vụ sang ngữ nghĩa(Enlightenment of services):**
 - a. Kiến trúc Lưới Ngữ Nghĩa nên giảm thiểu nhưng ảnh hưởng khi thêm tính ngữ nghĩa vào trong các thực thể , dịch vụ của Lưới thông thường. Nhằm giúp việc chuyển từ grid thường sang grid ngữ nghĩa được dễ dàng và nhanh chóng.
 - b. Các thực thể Lưới không bị phá vỡ nếu nó sử dụng và xử lý các tài nguyên Lưới nhưng xử lý các ngữ nghĩa kết hợp với các tài nguyên đó.
 - c. Trong suốt thời gian sống (lifetime) của chúng ,các thực thể Lưới có thể có hoặc mất tính ngữ nghĩa.

S-OGSA gồm có ba phần chính: mô hình (model, các thực thể Grid ngữ nghĩa), khả năng (capabilities, chức năng của các thực thể ngữ nghĩa), cơ chế (mechanism).

3. Mô hình và khả năng của S-OGSA

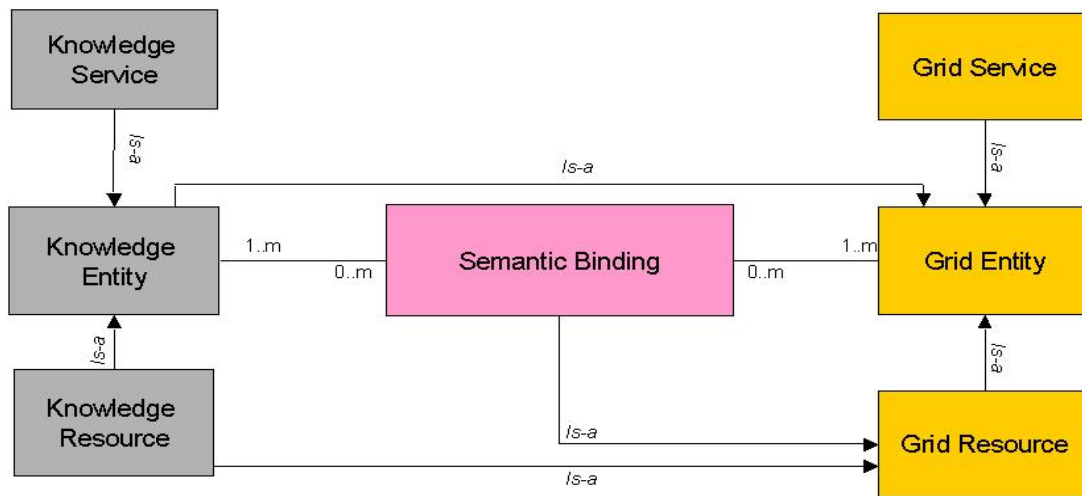
3.1 Mô hình S-OGSA

Mô hình của Lưới Ngữ Nghĩa được bao gồm các thành phần sau:

- a. Các thực thể Lưới (G-Entities): là bất kỳ đối tượng nào có số nhận dạng (identity) trong Lưới, bao gồm các tài nguyên và các dịch vụ.
- b. Các thực thể tri thức (K-Entities): đây là loại thực thể đặc biệt có thể hoạt động với các dạng tri thức. Các thực thể tri thức gồm : Ontologies, Rules, Cơ sở tri thức ,hay các mô tả dạng văn bản của tri thức mà có thể chia sẻ được. Các dịch vụ tri thức là những dịch

vụ có thể hoạt động và truy cập tới các tài nguyên tri thức như các động cơ luật, bộ suy diễn tự động ...

- c. Các ràng buộc ngữ nghĩa (S-Bindings): là các thực thể dùng để biểu diễn sự kết hợp giữa một thực thể Lưới thông thường (không có tính ngữ nghĩa) với một hoặc nhiều thực thể Lưới ngữ nghĩa.
- d. Các thực thể Lưới Ngữ Nghĩa (SG-Entities): là các các ràng buộc ngữ nghĩa hay các thực thể tri thức.



Hình 9 Mối quan hệ giữa các thực thể Lưới thông thường và Lưới Ngữ Nghĩa

3.2 Khả năng của S-OGSA

Theo nguyên lý thiết kế về tính đa dạng (diversity), Lưới Ngữ Nghĩa bao gồm một tập các dịch vụ các cấp độ khác nhau về khả năng xử lý ngữ nghĩa. Các dịch vụ ngữ nghĩa được chia thành hai nhóm: nhóm dịch vụ cung cấp ngữ nghĩa (Semantic Provisioning Services -SPS) và nhóm dịch vụ hiểu ngữ nghĩa (Semantically Aware Grid Services -SAGS). Nhóm dịch vụ SPS chịu trách nhiệm cung cấp và quản lý sự kết hợp ngữ nghĩa cho các thực thể Grid thông thường. Nhóm dịch vụ SAGS là những dịch vụ Lưới mở rộng có thể xử lý ngữ nghĩa.

3.2.1 Các dịch vụ cung cấp ngữ nghĩa

Các dịch vụ cung cấp ngữ nghĩa là các dịch vụ cho phép tạo, lưu trữ , cập nhật, loại bỏ và truy cập các dạng siêu dữ liệu và tri thức khác nhau.

Các dịch vụ cung cấp ngữ nghĩa được phân thành hai loại chính : nhóm dịch vụ cung cấp tri thức và nhóm dịch vụ cung cấp ràng buộc ngữ nghĩa (xem Hình 3).

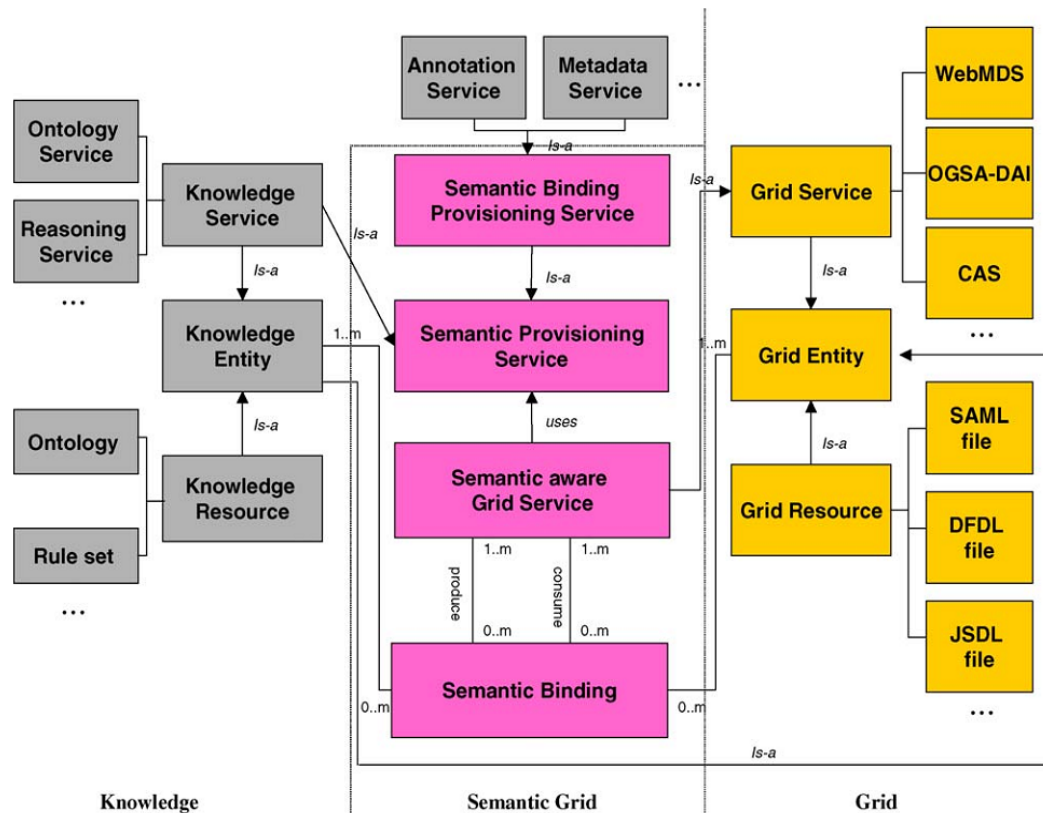
- a. **Nhóm dịch vụ cung cấp tri thức** (Knowledge provisioning services) gồm: các dịch vụ ontology chịu trách nhiệm lưu trữ ,truy xuất các mô hình tri thức ý niệm và các dịch vụ suy diễn chịu trách nhiệm tính suy diễn trên các mô hình tri thức này.

Các dịch vụ ontology cung cấp truy cập đến các tri thức lưu trữ trong các ontology, ở dạng các khái niệm và mối quan hệ giữa các khái niệm và ràng buộc của các mối quan hệ. Các ontologies có thể được truy cập thông qua các ngôn ngữ truy vấn hoặc các API.

Các dịch vụ suy diễn cho phép suy diễn các thông tin mới và kiểm tra các ràng buộc trên các tri thức có trong ontologies.

- b. **Nhóm Các dịch vụ cung cấp ràng buộc ngữ nghĩa gồm:** các dịch vụ về siêu dữ liệu (metadata services) chịu trách nhiệm lưu trữ truy cập đến các ràng buộc ngữ nghĩa, sinh ra các siêu dữ liệu cho các tài nguyên Grid từ nhiều nguồn thông tin khác nhau.

Dịch vụ siêu dữ liệu chịu trách nhiệm lưu trữ và cung cấp truy cập đến các ràng buộc ngữ nghĩa. Cũng giống như các dịch vụ Ontology, các siêu dữ liệu có thể được truy cập thông qua ngôn ngữ truy vấn hoặc các API. Có mối quan hệ chặt chẽ giữa các dịch vụ siêu dữ liệu và ontology, bởi vì dữ liệu được lưu trữ bởi dịch vụ siêu dữ liệu thông thường được dựa trên các mô hình ý niệm được lưu trữ trong các ontology. Các dịch vụ siêu dữ liệu có thể sử dụng các dịch vụ ontology và suy diễn để suy diễn các dữ liệu mà chúng lưu trữ.



Hình 10 Mô hình thông tin của Lưới Ngữ Nghĩa

3.2.2 Các dịch vụ Lưới Ngữ Nghĩa

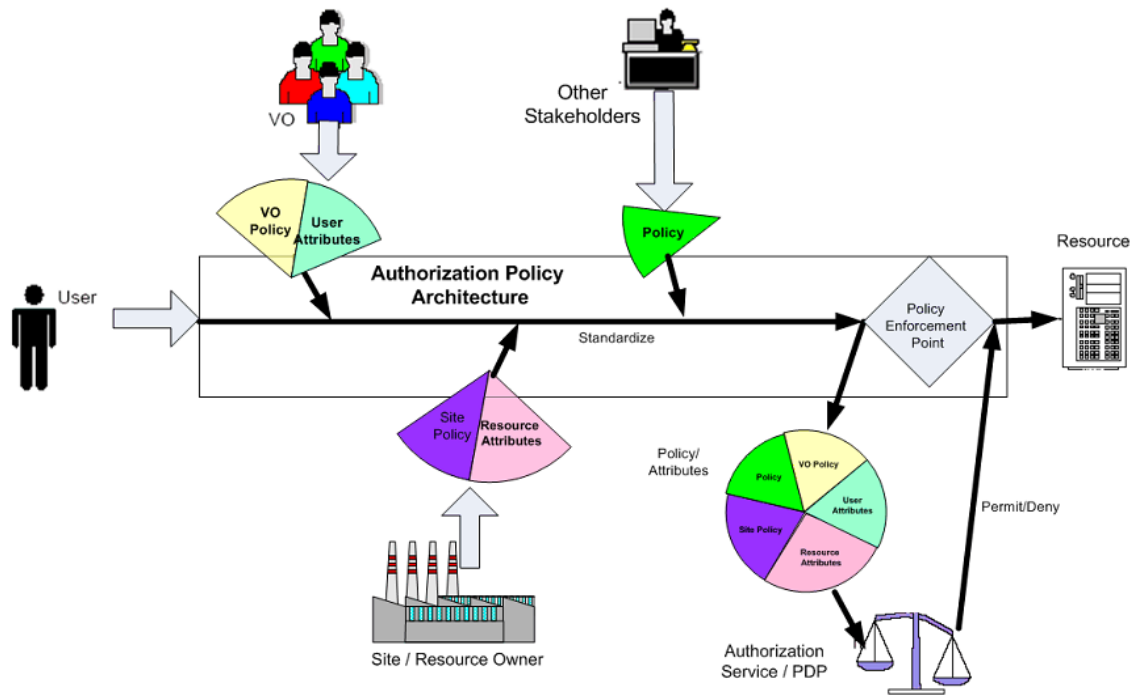
Là nhóm các dịch vụ có thể khai thác các kỹ thuật xử lý tri thức để tạo cung cấp các chức năng. Hiểu ngữ nghĩa tức là có khả năng sử dụng các ràng buộc ngữ nghĩa và hành động dựa trên các siêu dữ liệu và tri thức. Một số ví dụ về các hành động theo ngữ nghĩa như:

- Việc xác thực cho một đối tượng Grid dựa vào thông tin siêu dữ liệu được thực hiện bởi dịch vụ quản lý tổ chức ảo (VO Manager service).
- Việc thực thi một yêu cầu tìm kiếm trên các entries trong danh mục tài nguyên ngữ nghĩa.
- tích hợp một khái niệm mới vào ontology được quản lý bởi dịch vụ ontology.

Kiến trúc Lưới ngữ nghĩa cho phép việc chia sẻ tri thức giữa các cộng đồng Lưới một cách dễ dàng và thuận tiện.

3.3 Ví dụ về khả năng và mô hình S-OGSA trong ngữ cảnh hệ thống điều khiển truy xuất có ngữ nghĩa

Chúng ta xét ví dụ sau, tổ chức ảo cung cấp ngữ cảnh cho các thao tác của Lưới mà nó được sử dụng để kết hợp người sử dụng, các yêu cầu của họ và một tập các tài nguyên. Các nhà cung cấp tài nguyên và người tiêu dùng định nghĩa một cách rõ ràng và cẩn thận những gì được chia sẻ, ai được phép chia sẻ và dưới những điều kiện nào thì cái gì được chia sẻ... Mỗi tổ chức mà tham gia vào tổ chức ảo (bao gồm các viện, cá nhân, các nhóm, ...) định nghĩa một tập các thiết yếu chính sách bao gồm các luật truy xuất vào nguồn tài nguyên của họ. Các tập tin chính sách được định nghĩa một cách độc lập và được phân tán nhiều vùng khác nhau trong tổ chức ảo, như được hiển thị trong hình sau đây.



Hình 11 Kiến trúc chính sách xác thực quyền

Khi một người sử dụng muốn truy cập một dịch vụ trong tổ chức ảo, nó sẽ liên lạc với dịch vụ mà điều khiển việc truy xuất tới những tài nguyên (là Policy Enforcement Point trong hình). Dịch vụ này điều hòa các chính sách khác nhau để mà Policy Decision Point (authorization service/PDP trong hình) có thể giao tiếp với chúng và quyết định người sử dụng này có quyền truy cập hay không.

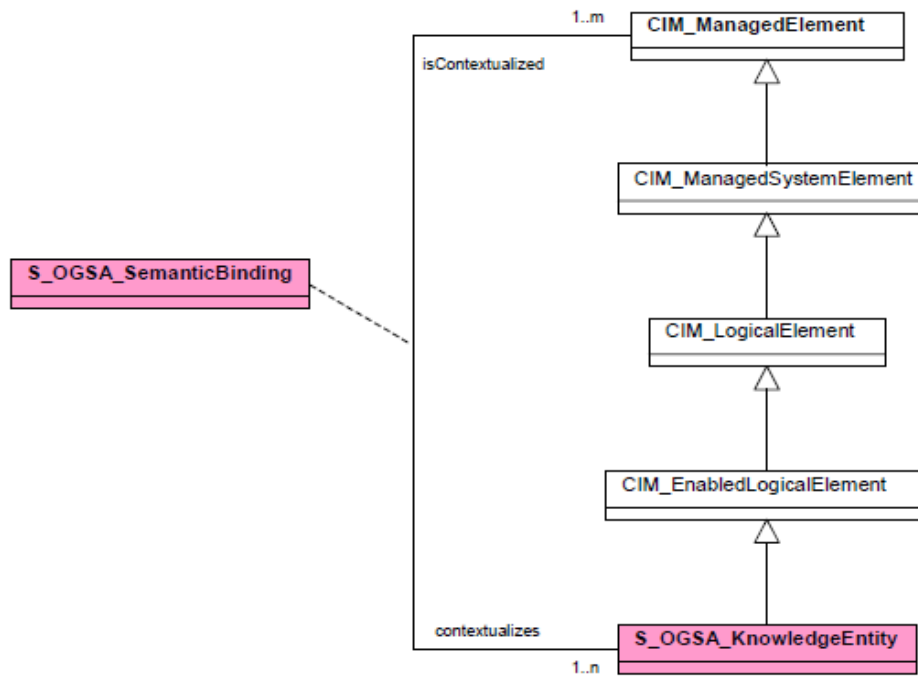
4. Cơ chế của S-OGSA

4.1 Đối xử các thực thể tri thức và việc kết hợp ngữ nghĩa như là những tài nguyên của lưới

Nhóm nghiên cứu đã chọn mô hình tài nguyên Common Information Model (CIM) để hiện thực những phần mở rộng của S-OGSA đối với Knowledge Entities và Semantic Binding. CIM là mô hình hướng đối tượng, độc lập về kỹ thuật cho việc định nghĩa các đối tượng được quản lý trong thế giới thực trong môi trường máy tính và mạng máy tính. CIM đã được được phát minh bởi tổ chức Distributed Management Task Force (DMTF).

Mô hình CIM hiện tại được biểu diễn bằng UML, mặc dù các dạng biểu diễn khác cũng đang được phát triển. Nhóm nghiên cứu đã minh họa cho phát biểu của họ bằng cách hiện thực các thực thể cơ bản của S-OGSA như là phần mở rộng của mô hình tài nguyên CIM trong UML (tham khảo Hình 4).

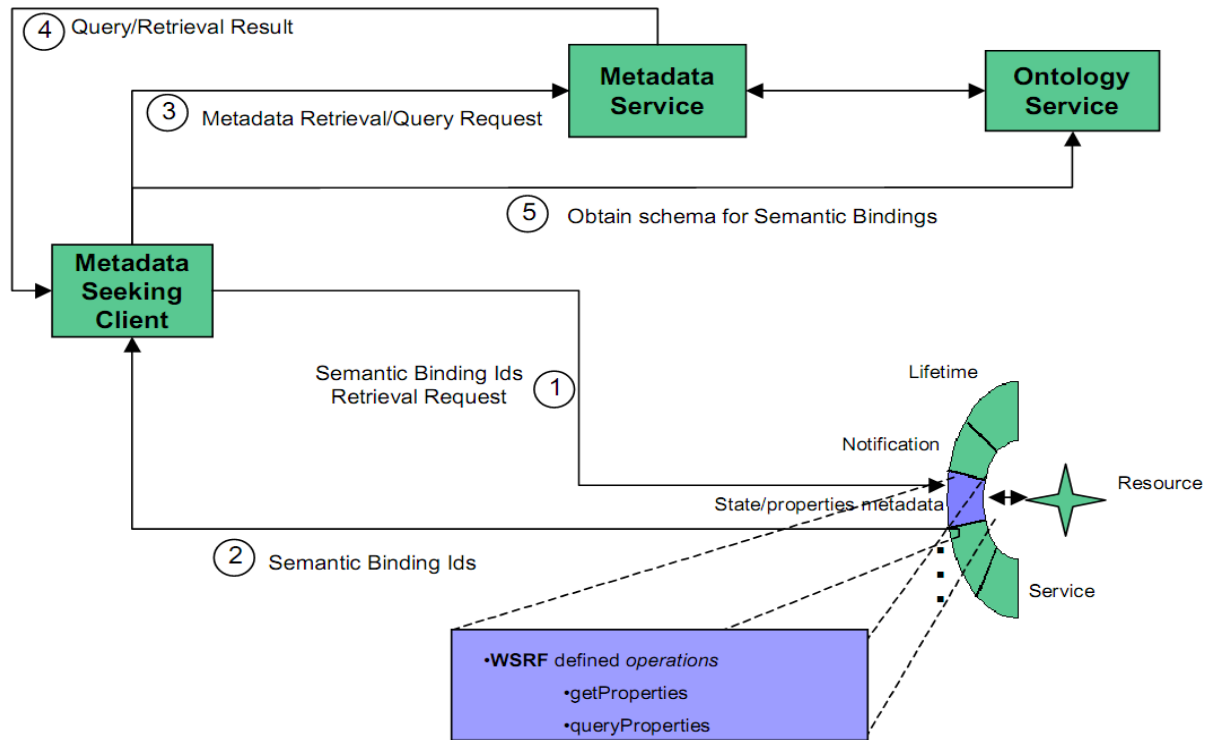
- a. Grid Entities được biểu diễn bằng lớp CIM-ManagementElement trong mô hình CIM.
- b. Knowledge Entities được biểu diễn bằng lớp mới S-OGSA-KnowledgeEntity, lớp con trực tiếp của lớp CIM-ManagedElement.
- c. Cuối cùng, việc kết hợp giữa một Grid Entity (CIM-ManagedElement) và Knowledge Entity (S-OGSA-KnowledgeEntity) được biểu diễn bằng lớp mới S-OGSA-SemanticBinding.



Hình 12 Mô hình S-OGSA như phần mở rộng của mô hình CIM

4.2 Các dịch vụ S-Stateful

Nhóm nghiên cứu định nghĩa các dịch vụ S-Stateful là những dịch vụ mà các tài nguyên Lưới ảo hóa được kết hợp với nhau cùng với siêu dữ liệu. Đặc tả của dịch vụ S-Stateful là một tập các ánh xạ cùng với mô hình S-OGSA để hiện thực cho Lưới hướng dịch vụ. Dịch vụ S-Stateful là cơ chế phân phát.



Hình 13 Truy xuất và truy vấn ràng buộc ngữ nghĩa của tài nguyên

Hình 6 mô tả một dạng có thể của việc tương tác đối với việc phân phát siêu dữ liệu được mã hóa có ngữ nghĩa, chúng ta có thể tóm tắt như sau:

- Metadata (Semantic Bindings) được xem như nguồn tài nguyên, được quản lý thông qua dịch vụ siêu dữ liệu, như trong hình minh họa. Các dịch vụ cung cấp việc truy xuất tới các tài nguyên cùng với các thuộc tính được tạo tương ứng với các kiểu dữ liệu lược đồ XML được sinh ra bởi quá trình ánh xạ.
- Khi có yêu cầu, các tài nguyên Lưới có thể cung cấp các tài liệu tham khảo điểm cuối của các tài nguyên có ràng buộc ngữ nghĩa được kết hợp với chúng thông qua các hoạt động phân phối siêu dữ liệu cụ thể ở mức cơ sở hạ tầng.
- Các khách hàng quan tâm tới việc khai thác Ràng buộc ngữ nghĩa có thể tương tác với các dịch vụ siêu dữ liệu thông qua bất kỳ hoạt động nào mà tài nguyên Ràng buộc ngữ nghĩa hỗ trợ.

III. ỨNG DỤNG GRID NGŨ NGHĨA MYEXPERIMENT

Các myExperiment Virtual nghiên cứu môi trường hỗ trợ việc chia sẻ của các đối tượng nghiên cứu cho các nhà khoa học sử dụng, chẳng hạn như các công việc khoa học. Đối với các nhà nghiên cứu nó là cả một cơ sở hạ tầng xã hội khuyến khích chia sẻ và là nền tảng để thực hiện nghiên cứu thông qua giao diện người dùng quen thuộc. Chúng cung cấp một môi trường mở rộng cho phát triển, mô tả việc thiết kế, thực hiện và triển khai các myExperiment và có

bốn khả năng chính: đối tượng nghiên cứu, mô hình xã hội, môi trường mở và hoạt động nghiên cứu, đây là những đặc tính cần thiết của kết quả nghiên cứu môi trường điện tử và khoa học mở.

Tiền bộ khoa học dựa vào một quá trình xã hội mà các nhà khoa học chia sẻ ý tưởng, phương pháp và dữ liệu. Theo truyền thống, thuyết này là trung gian của quá trình xuất bản, nhưng các nhà khoa học dần chuyển sang các blog, wiki và mạng xã hội để tạo thuận lợi cho quá trình này, một hiện tượng đôi khi mô tả như là Khoa học 2.0 [1]. Với các dạng này chúng ta thấy mở ra một phong trào khoa học với quy mô lớn, mở ra sự kết hợp, chia sẻ dữ liệu trên các trang Web. Mục đích của việc nghiên cứu môi trường ảo(VRE) là "cung cấp cho các nhà nghiên cứu các công cụ và dịch vụ mà họ cần để làm nghiên cứu bất kỳ lúc nào để mang lại hiệu quả nhất có thể "[2]. Phản ánh quan sát của chúng ta vào quá trình phát triển của khoa học, chúng cho thấy kết quả nghiên cứu môi trường ảo nên cung cấp bốn khả năng quan trọng, và chúng ta đề xuất là những định nghĩa của " Nghiên cứu Môi trường Xã hội Ảo ": Nó sẽ tạo thuận lợi cho việc quản lý và chia sẻ Đối tượng nghiên cứu - đó là những kỹ thuật số hàng hóa được sử dụng và tái sử dụng bởi các nhà nghiên cứu, từ những dữ liệu và phương pháp xuất bản tác phẩm học thuật.

Nó sẽ hỗ trợ các mô hình xã hội: nhà sản xuất các đối tượng nghiên cứu nên có ưu đãi để; người tiêu dùng cần để có thể khám phá và tái sử dụng chúng, tất cả sẽ được hưởng lợi từ cộng đồng.

Nó sẽ cung cấp một môi trường mở rộng cho phép dễ dàng tích hợp với phần mềm khác bằng các công cụ và dịch vụ, và hưởng lợi từ sự tham gia đóng góp của phần mềm.

Nó sẽ cung cấp một nền tảng để nghiên cứu hành động. Ví dụ để cung cấp cho các đối tượng nghiên cứu dịch vụ từ xa và phần mềm. Nó sẽ được đơn giản để tạo ra các tùy chỉnh, các công cụ một môi trường công việc cụ thể. Khả năng (1) là một chức năng lưu trữ, (2) là đặc trưng của các trang web xã hội và (3) có liên quan đến phát triển cộng đồng mã nguồn mở, (4) là một môi trường nghiên cứu - nghiên cứu các đối tượng không chỉ là lưu trữ và trao đổi mà còn được sử dụng trong việc tiến hành nghiên cứu (chúng ta mô tả chúng như là hành động). Chúng ta lưu ý rằng các nguyên lý của khoa học mở - dữ liệu mở, truy cập mở và mã nguồn mở - là phù hợp với định nghĩa này. Tiềm ẩn trong tất cả các khả năng là quan điểm cho rằng giao diện phải thân thiện để con người dễ sử dụng và làm quen. Nhằm thiết lập bốn khả năng chúng ta đã thiết kế và xây dựng myExperiment, một trang web dành cho các nhà khoa học trực tiếp hỗ trợ nghiên cứu của họ. Các dịch vụ myExperiment.org ra đời trong tháng đã 11 năm 2007 và đã thu hút được sự quan tâm đáng kể. Trong thời gian bảy tháng đầu năm 2008, trang web nhận được trên 8.500 du khách và duy nhất đạt được trên 1.000 người dùng đăng ký. myExperiment là đặc biệt bởi vì :

1. Nó chuyên sâu vào các khía cạnh xã hội, và nó có thể được xem như là một thử nghiệm phá khoa học để cộng đồng chia sẻ và để được hưởng lợi từ mạng lưới các hiệu ứng của một trang web xã hội. Trong bài báo này chúng tôi báo cáo lần đầu tiên trên việc xây

dụng và sử dụng các trang web, và những hiểu biết trong bốn khả năng. Phần 2 trình bày các dự án myExperiment trong khuôn khổ các khả năng và vị trí của nó đối với công việc liên quan. Chúng tôi sau đó nhìn vào phần mềm thiết kế tại Mục 3 triển khai và thực hiện, ở mục 4. Sau khi phân tích về cách sử dụng. Mục 5, chúng tôi đóng, ở Phần 6 xét lại bốn khả năng và kinh nghiệm của chúng ta về "các myExperiment thử nghiệm" ở giai đoạn mình phát triển.

2. Các myExperiment VRE Khoa học công việc là hàng hóa có giá trị trong đó yêu cầu chuyên môn để xây dựng [3]. myExperiment được thúc đẩy bằng cách quan sát một nhu cầu rõ ràng để chia sẻ quy trình công việc - để làm giảm sự tái tạo, tuyên truyền tốt nhất thực hành và cho phép các nhà khoa học phải tập trung vào khoa học - Giữa một cộng đồng khá tách rời của công việc người sử dụng. Nó cũng được thúc đẩy bởi một sự thất vọng với các hệ thống đó: (a) bị mất việc xã hội kích thích, chỉ làm hứng thú có sẵn thay vì khuyến khích và kiểm soát chia sẻ, (b) trình bày phức tạp trong giao diện người dùng phù hợp với phổ biến các trang web mà người đang sử dụng trên cơ sở hàng ngày, do đó đòi hỏi kỹ năng hơn nữa. Các động lực và lý do cho các dự án myExperiment được thảo luận trong chi tiết trong [4-5] và các nguyên tắc thiết kế trong [6].

2.1. Khả năng myExperiment chúng tôi định nghĩa như sau:

2.1.1. Đối tượng nghiên cứu chính của chúng tôi là các đối tượng khoa học và đối tượng của liên quan (chẳng hạn như dữ liệu và tài liệu). Chúng tôi có hỗ trợ thêm cho định dạng công việc cụ thể để chúng tôi có thể 'nhìn vào bên trong' các đối tượng này để trích xuất các hợp chất siêu dữ liệu, cung cấp đồ họa và có thể xác định các dịch vụ được sử dụng. Chúng tôi đã cung cấp đầy đủ các hỗ trợ cho công việc Taverna [7] và chúng tôi hiện đang phát triển hỗ trợ cho các hệ thống khác. MyExperiment đáng kể cũng hỗ trợ nghiên cứu các đối tượng đó là bộ sưu tập của các đối tượng khác, vì các nhà nghiên cứu làm việc với các bộ sưu tập các mặt hàng liên quan với một thử nghiệm - ví dụ, một phiên bản cụ thể một quy trình làm việc cùng với đầu vào và đầu ra dữ liệu, dịch vụ gọi các bản ghi và tài liệu. Họ có thể cũng thu thập nhiều công việc với nhau để chia sẻ. Những bộ sưu tập được thể hiện cho người dùng như các gói và được một Tính năng đặc biệt của myExperiment. Khi các cộng đồng người sử dụng tăng myExperiment về số lượng và bề rộng, chúng tôi đang phát triển hỗ trợ mới nghiên cứu các đối tượng, chẳng hạn như kế hoạch thử nghiệm và các mô hình thống kê.

2.1.2. Mô hình xã hội: Để hỗ trợ sản xuất của nghiên cứu đối tượng đóng góp cho myExperiment chúng tôi cung cấp các thành viên của trang web với hỗ trợ tín dụng và ghi công, và kiểm soát tốt và khả năng hiển thị chia sẻ của các đối tượng nghiên cứu. Đầu người sử dụng thông tin phản hồi tiết lộ này là yếu tố quan trọng nhất trong việc đưa ra một

trang web xã hội chấp nhận được để sử dụng bởi các nhà khoa học. Các thành viên khác của "tiêu thụ" các trang web nghiên cứu các đối tượng có thể xem, tải về, tag, xem xét và 'Ưu thích' họ, trong đó viện trợ phát hiện và tăng cường uy tín của các nhà sản xuất. Ngoài ra, nội dung tiếp xúc công khai được phát hiện thông qua công cụ tìm kiếm. Trừ khi họ được duy trì, công việc và khác các đối tượng nghiên cứu có thể chấm dứt được tái sử dụng theo thời gian - họ có hiệu quả "phân rã", mặc dù trong thực tế nó là của mình bởi cảnh đó đang thay đổi. Ví dụ, một thay đổi gần đây trong nhận dạng gen của một nhà cung cấp dịch vụ đã dẫn đến một myExperiment thông báo cho người sử dụng bị ảnh hưởng quy trình công việc. công việc hữu ích sẽ được giám tuyển của cộng đồng sử dụng chúng, và các tác giả gốc Cha phụ tá cũng được khuyến khích bởi vì họ đang nhận được tín dụng sử dụng tác phẩm của họ. Quy trình làm việc sâu rộng là một vấn đề khó khăn và myExperiment cung cấp một mới phương pháp tiếp cận thông qua curation cộng đồng.

2.1.3 Môi trường mở. myExperiment đã tạo nhiều sự chú ý cho cộng đồng phát triển của nó như là nó phải thiết kế giao diện người dùng. Bằng cách tạo ra các công cụ để quản lý các API, các tiếp xúc. Chức năng tùy biến cao để đáp ứng yêu cầu. Các API đã được kích hoạt giao diện mới được xây dựng, chẳng hạn như Google Gadgets và Facebook Apps. Nó cũng cho phép các giao diện hiện tại để kết hợp chức năng myExperiment, như một wiki hoặc Taverna công việc bàn làm việc. myExperiment luôn thích sử dụng lại để tái tạo và có thể dễ dàng truy cập các dịch vụ khác. Nó được thiết kế là một phần của chu kỳ kiến thức học thuật và là tương thích với các giao thức Open Archives Initiative. Trong khi nó cung cấp một chức năng công việc lưu trữ, nhiều thông tin liên quan - chẳng hạn như dữ liệu và ấn phẩm - có thể được tổ chức tại các kho khác, do đó, myExperiment làm cho nó dễ dàng để tham khảo bên ngoài nội dung.

Ngoài các API, các cơ sở mã myExperiment là mã nguồn mở và có thể được sử dụng bởi bất cứ ai để thiết lập myExperiment dự của riêng mình.

2.1.4 Hoạt động nghiên cứu myExperiment được thiết kế kêu gọi các dịch vụ bên ngoài quá trình nghiên cứu các đối tượng. Taverna công việc được thực hiện bởi myExperiment nộp một bộ sưu tập của nghiên cứu các đối tượng để xử lý từ xa để enact một, và kết quả sẽ được tự động thu lại vào myExperiment. Một cơ chế tương tự có thể chạy mô phỏng hoặc các mô hình thống kê, ví dụ. Các dịch vụ có thể được địa phương để người sử dụng, có lẽ trong phòng thí nghiệm của họ, hoặc

có tiềm năng " trong đám mây". Điều này khả năng thứ hai là quan trọng bởi vì các nhà nghiên cứu sau đó có thể truy cập dịch vụ từ xa mà không cần bất kỳ yêu cầu để cài đặt phần mềm địa phương. Cũng như mang lại khả năng này cho người dùng thông qua giao diện myExperiment, API được thiết kế sao cho các nhà phát triển có thể dễ dàng xây dựng 'chức năng mashup' trên myExperiment cho nhanh tạo mẫu của các công cụ để hỗ trợ các nhà nghiên cứu. Đây có thể được giao diện quy tắc cho các nhiệm vụ cụ thể, chẳng hạn như chạy công việc cấu hình sẵn.

3. Thiết kế phần mềm

myExperiment được thiết kế xung quanh một tập hợp các thực thể được phản ánh trong các mô hình dữ liệu nội bộ, người sử dụng giao diện (xem hình 1) và dữ liệu mở bên ngoài đại diện. Đây là nguồn gốc thông qua mở rộng người dùng tham khảo ý kiến với các nhóm tập trung và giao diện mock-up.

3.1. Các đối tượng trong myExperiment đang được mở rộng để hỗ trợ nhiều đối tượng nghiên cứu trong các lĩnh vực khác nhau. Các đối tượng nghiên cứu hiện nay là:

Quy trình công việc - hợp chất có chứa đối tượng dịch vụ, các biểu đồ công việc, công việc cụ thể siêu dữ liệu và thông tin bổ sung phụ thuộc vào hệ thống công việc. Quy trình công việc được versioned, và mỗi công việc thống kê sử dụng như số lần xem của và số lượng tải về. · Gói - bộ sưu tập của các đối tượng nghiên cứu để hình thành tổng hợp các thực thể. Ngoài các đối tượng trên hiện tại máy chủ, bao cũng có thể chứa các liên kết đến các đối tượng trên các máy chủ khác. · Tập tin - đối tượng nhị phân được tải lên myExperiment và được đọc vào hệ thống. · Groups - bộ sưu tập của người sử dụng đăng ký. Các Người tạo ra một nhóm điều khiển của nó thành viên thông qua các lời mời và yêu cầu. Các mô hình dữ liệu hỗ trợ các mối quan hệ giữa các nhóm.

The screenshot displays the myExperiment website interface. At the top, there's a navigation bar with links for Home, Users, Groups, Workflows, Files, and Packs. A search bar is also present. The main content area shows a workflow entry titled 'Retrieve Protein Sequence (Taverna Workflow)'. It includes metadata such as creation and update dates, and a list of statistics (Licence, Credits, Attributions, Tags, etc.). The workflow diagram is visible, showing inputs like 'ID' and 'database', and outputs like 'searchSimple' and 'Get_Protein_FASTA'. On the right side, there's a sidebar with user information for David De Roure, including a profile picture, name, and links to My Profile, My Messages, My Memberships, My History, and My News. Below this, there's a 'My Stuff' section showing 23 friends, 5 groups, 1 file, and 1 workflow. The 'Friends' list includes names like Andrea Viggins, Anton Godens, Carole Goble, Catherine De Roure, Darius Michaelides, and David De Roure 3.0.

Mô hình xã hội được hỗ trợ thông qua một số thực thể.

Trước hết là các thành viên - đây là những người sử dụng đã đăng ký với myExperiment và có thể tìm thấy hoặc đóng góp các đối tượng nghiên cứu, tạo ra các thẻ, nhận xét và đánh giá, vv Họ cũng có thể kết bạn với các thành viên. Một người sử dụng rất nhiều các myExperiment không thành viên - họ chỉ đơn giản là người duyệt trang web công bố công khai nội dung hoặc liên kết sau đây từ công cụ tìm kiếm như Google. 'siêu dữ liệu xã hội' Các đơn vị khác, trong đó nhiều có thể nhìn thấy trong hình 2, là: 1. Attributions - Vì vậy mà các thành viên có thể hiển thị những gì một đối tượng nghiên cứu được dựa trên Creditations - Vì vậy mà các thành viên có thể cung cấp tín dụng những người khác. Theo mặc định, người tải lên là được tín dụng. Ảnh - Để hỗ trợ và cung cấp danh tiếng khuyến khích, các thành viên có thể xác định yêu thích của họ nghiên cứu đối tượng. Danh sách yêu thích được nhìn thấy các thành viên khác.

4. Đánh giá - A 5 sao hệ thống đánh giá đơn giản để hỗ trợ với các khuyến nghị.
5. Nhận xét - Giải thích để gia tăng xếp hạng.
6. Trích dẫn - Xuất bản thông tin liên quan nghiên cứu đối tượng.
7. Bình luận - Để cho phép các thành viên để nhận xét về nghiên cứu đối tượng.

8. Tags - Để chú thích các đối tượng nghiên cứu để dễ khám phá. Chủ sở hữu của "gắn thẻ" (các Hiệp hội các tag để đối tượng) được ghi lại.
9. Chính sách - xem dưới đây để sở hữu, chia sẻ và Cho phép. Nói chung giao diện người dùng phản ánh cùng một bộ các đơn vị và được thiết kế để làm cho nó dễ dàng như thể cho người tiêu dùng để tìm các đối tượng nghiên cứu (bằng cách tìm kiếm hoặc chuyển hướng) và các nhà sản xuất để đóng góp. Công việc, người dùng, nhóm và các gói có của mình các trang web mà trở thành chủ đề pivoting và duyệt. Các cơ chế gắn thẻ và cho ý kiến vv là nhất quán trên các trang này. Chỉ nội dung có quyền được hiển thị có thể nhìn thấy cho người sử dụng hiện hành. Kết quả là, nhiều bộ phận của giao diện người dùng rất năng động, với các nội dung khác nhau, tính năng và hành động hiển thị / kích hoạt dựa trên bối cảnh người dùng hiện hành.



Uploader

Marco Roos

License

All versions of this Workflow are licensed under the **Creative Commons Attribution-Share Alike 3.0 License**.

Credits (2)

(People/Groups)

Marco Roos

AID

Attributions (1)

(Workflows/Files)

BioAID_DiseaseDiscovery

Tags (8)

All Tags

Uploader Tags

AIDA | BioAID | biorange_nl | count | disease | text_mining | text_mining_network | VL-e

Add tags

Shared with Groups (1)

AID

Featured In Packs (0)

None

Add to your Pack

Ratings (0)

Hover and click to rate

Current: 5.00/5 (1 rating)

You rated it: 5/5

Breakdown

Attributed By (0)

(Workflows/Files)

None

Favourited By (0)

No one

Add to your Favourites

Statistics

783 viewings
652 downloads
2 credits
1 attribution
8 tags
0 favourites
0 ratings
0 citations
0 reviews
0 comments

Đôi khi các giao diện người dùng không trực tiếp phản ánh các mô hình dữ liệu cơ bản. Trong đó, chia sẻ và quyền cơ bản của mô hình được đánh giá cao hướng đối tượng, với người dùng, nhóm, đóng góp, chính sách và các đối tượng Giấy phép tất cả làm việc cùng nhau. Tuy nhiên, người sử dụng được trình bày với 'đóng hộp' tùy chọn cho phép nhanh chóng lựa chọn các tùy chọn chia sẻ thích hợp nhất, ví dụ: "bất cứ ai có thể xem và tải "; "bất cứ ai có thể xem, nhưng chỉ bạn bè của tôi được phép download ".

Chúng ta phát triển một phần mềm myExperiment tính năng thường xuyên yêu cầu là khả năng tải lên các nội dung khác ngoài các công việc và sau đó liên kết với các phần khác nhau để cho một mục đích cụ thể. Điều này phản ánh thực tế là các nhà khoa học làm việc với các bộ sưu tập nghiên cứu các đối tượng, chẳng hạn như dữ liệu đầu vào và đầu ra cho một công việc hoặc một tập hợp các quy trình công việc. Do đó chúng tôi đặt ra một khái niệm tổng quát hơn nghiên cứu đối tượng, cũng bao gồm các hình thức khác của dữ liệu trong myExperiment - cho các thành viên. Ví dụ nhóm, thẻ và các mạng xã hội. Chúng tôi kêu gọi các đối tượng emos (Đối tượng myExperiment Encapsulated). Emos được đại diện trong RDF và chúng tôi đã phát triển một myExperiment trong đó sử dụng Dublin Core siêu dữ liệu cho các đối tượng nghiên cứu và FOAF cho mạng lưới thông tin. Để đáp ứng các yêu cầu của các nhà khoa học, emos mang thông tin cho phép nội dung có thể thay đổi để được xác nhận. Emos có thể xuất ra định dạng để hỗ trợ hội nhập. Để làm việc liên quan với kho dữ liệu chúng ta có thể thông qua các đại diện đối tượng tái sử dụng và chuyển đổi từ các sáng kiến Lưu trữ Mở (www.openarchives.org / quặng) dựa trên tên đồ thị RDF.

IV. PHỤ LỤC PHÂN CÔNG CÔNG VIỆC

Đây là bài báo cáo chi tiết của seminar 5 với chủ đề tài “*Semantic Grid: From Concepts to Implementation*” với công việc cụ thể như sau:

Nội dung	Người thực hiện
I.Grid ngữ nghĩa	Phạm Thị Bích Liên
II.Kiến trúc Grid ngữ nghĩa S-OGSA	Nguyễn Văn Nhất
III.Ứng dụng grid ngữ nghĩa: MyExperiment	Nguyễn Duy Thiện