

Thinking Big: Web Scale AI Thinking Big: Web Scale AI

Michael Witbrock

Cycorp Europe

witbrock@cycorp.eu

March 13th 2009

Michael Witbrock

Cycorp Europe

witbrock@cycorp.eu

March 13th 2009



Tutorial



OpenCyc Web Services

Looking up Semantic Terms for English Words

- ...
- warplanes
 - bombers (warplanes)
 - B 24 bomber
 - B-1 bombers
 - B-2 stealth bombers
 - B-29 Superfortress
 - B-52 bombers

```

Mt : \(ContextOfPCWFn MilitaryThingTaxonomy\)
broaderThan : ● Warplane

Mt : TaxonomyMt
nodeInSystem : ● MilitaryThingTaxonomy ● WeaponTaxonomy

Mt: EnglishMt
● \(multiWordString
  \(TheList "bomber"\) Aircraft-TheWord CountNoun BomberPlane\) ✖

Mt: GeneralEnglishMt
● M\(denotation Bomber-TheWord CountNoun 0 BomberPlane\) ✖
● \(genStringAssertion BomberPlane
  ● M\(denotation Bomber-TheWord CountNoun 0 BomberPlane\)\) ✖

```

As of Feb. 24, Air Force ([#\\$UnitedStatesAirForce](#)) officials reported that personnel in the area numbered close to 8,000. The 100 aircraft based in Saudi Arabia for patrols over southern Iraq has seen the addition of two dozen F-15 ([#\\$FighterPlane-F15](#)) and F-16 fighter jets ([#\\$FighterPlane-F16](#)) to Bahrain. The Air Force ([#\\$UnitedStatesAirForce](#)) has also authorized the dispatch of 12 F-117 ([#\\$FighterPlane-F117](#)) stealth fighter jets to Kuwait, three B-1 bombers ([#\\$B-1-Bomber](#)) to Bahrain and 14 B-52 ([#\\$B-52-Bomber](#)) bombers ([#\\$BomberPlane](#)) to the island of Diego Garcia. It also has diverted dozens of support aircraft to the region for refueling, surveillance, electronic jamming and search-and-rescue missions.

```

Mt : TaxonomyMt
nodeInSystem : ● MilitaryThingTaxonomy ● WeaponTaxonomy
nodesForConceptFilter : ● \(ConceptNodesForTaxonFn MilitaryThingTaxonomy\)

```



Precise Conceptual Tagging

- **Look up terms in OpenCyc**
- Generalise them
- Web Service: Find

Query:

<http://ws.opencyc.org/webservices/concept/find?str=dog&searchType=cycl&exactMatch=true&ignoreCase=true&uriType=readable>

Response:

```
<concepts xmlns="http://ws.opencyc.org/xsd/CycConcepts" hasMore="false">  
  <concept uri="http://sw.opencyc.org/2008/06/10/concept/en/Dog"/>  
</concepts>
```

Xpath: /concepts/concept/@uri

Value: <http://sw.opencyc.org/2008/06/10/concept/en/Dog>



Looking up terms

- XML documentation for find can be found at:
<http://ws.opencyc.org/wadl/OpenCycWS.wadl>
- XML Schema for the returned data:
<http://ws.opencyc.org/xsd/CycConcepts.xsd>

```

<!--
  Document   : find.xml
  Created on : September 17, 2008, 3:47 PM
  Author      : tbrussea
  Description:
    This is a WADL file that describes the "http://ws.opencyc.org/services/find/"
    webservice. See: https://wadl.dev.java.net/ for more details on WADL.
-->

<application
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://research.sun.com/wadl/2006/10 https://wadl.dev.java.net/wadl20061109.xsd"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns:ocyc="http://ws.opencyc.org/xsd/CycConcepts"
  xmlns="http://research.sun.com/wadl/2006/10">
  <grammars>
    <include href="http://ws.opencyc.org/xsd/CycConcepts.xsd"/>
  </grammars>
  <resources base="http://ws.opencyc.org/">
    <resource path="webservices/concept/find">
      <method name="GET" id="find">
        <request>
          <param name="str" type="xsd:string" style="query" required="true">
            <doc xml:lang="en" title="str (Sample Value: trom)">
              A string (possibly a partial string) representation of a Cyc
              concept for which more details need to be found.
            </doc>
          </param>
          <param name="maxResults" type="xsd:positiveInteger" style="query" required="false" default="100">
            <doc xml:lang="en" title="maxResults (Sample Value: 100)">
              The maximum number of results that this query should return.
              Note, there may be server specific limits that cause
              results to be truncated earlier than specified. A value of 0
              or less for this paramter will result in an error.
            </doc>
          </param>
        </request>
      </method>
    </resource>
  </resources>
</application>

```

WADL



But first: Documentation

```

public CycConceptsForString(String searchTerm) {
    if (searchTerm == null) {
        throw new RuntimeException("CycConceptsForString got null input");
    }
    theQuery = searchTerm;
    if (theQuery.length() < 3) return;

    try {
        String encSearch = URLEncoder.encode(searchTerm, "UTF-8");
        URI nsURI = new URI("urn:opencyc:ws");

```

```

        URI address = new URI("http", null, "ws.opencyc.org", 80,
            "/webservices/concept/find",
            "str=" + encSearch + "&isExactMatch=true", null);
        service.addPort(portName, HTTPBinding.HTTP_BINDING, address.toString());
        Dispatch<Source> dispatcher = service.createDispatch(portName,
            Source.class, Service.Mode.PAYLOAD);

```

```

            Source.class, Service.Mode.PAYLOAD);
        Map<String, Object> requestContext = dispatcher.getRequestContext();
        requestContext.put(MessageContext.HTTP_REQUEST_METHOD, new String("GET"));
        StreamSource result = (StreamSource) dispatcher.invoke(null);

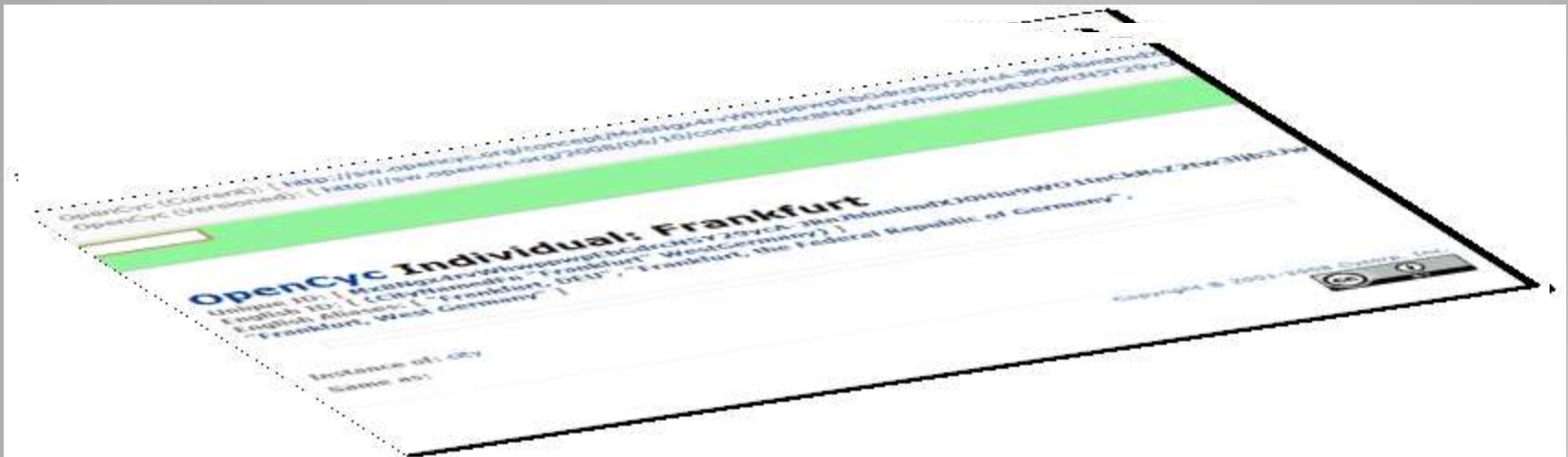
        if (result == null) {
            if (debug) System.out.println(address + " failed to get response for "
                + encSearch + " in CycConceptsForString");
            return;
        }
        InputStream inStream = result.getInputStream();

```

The code is calling the service

<http://ws.opencyc.org/webservices/concept/find?str=Frankfurt&isExactMatch=true>

```
<concepts xmlns="http://ws.opencyc.org/xsd/CycConcepts" hasMore="false">
  <concept uri="http://sw.opencyc.org/concept/Mx4rvdn2TZwpEbGdrcN5Y29ycA"/>
  <concept uri="http://sw.opencyc.org/concept/Mx4rvuusEpwpEbGdrcN5Y29ycA"/>
  <concept uri="http://sw.opencyc.org/concept/Mx8Ngh4rvYkdfpwpEbGdrcN5Y29ycB4rvdn2TZwpEbGdrcN5Y29ycA"/>
  <concept uri="http://sw.opencyc.org/concept/Mx8Ngx4rvWhwppwpEbGdrcN5Y29ycA-JRnJhbmtmdXJ0Hiu9WO1tnCkRs"/>
</concepts>
```



Manual Call to WS


```

Document response =
    DocumentBuilderFactory.newInstance().newDocumentBuilder().parse(inStream);

XPathFactory xpf = XPathFactory.newInstance();
XPath xp = xpf.newXPath();
NodeList resultList = (NodeList) xp.evaluate(
    "/concepts/concept", response, XPathConstants.NODESET);
int len = resultList.getLength();
for (int i = 0; i < len; i++) {
    String title = (String) xp.evaluate("@uri", resultList.item(i), XPathConstants.STRING);
    String conceptID = (title==null)
        ?null
        :URLDecoder.decode(title.replaceFirst(uriPrefix, ""), "UTF-8");
    if (concepts == null && conceptID != null) {
        concepts = new ArrayList<String>(defaultConceptCount);
    }
    concepts.add(conceptID);
}

```

Extracting the Concepts

NGrams: Asia-Europe, summit, begins, in, Beijing, amid, financial, crisis, Chinese, Premier, Wen, Jiabao, opened, the, two-day, Asia-Europe, Meeting, here, on, Friday, with, the, summit, attended, by, leaders, from, 43, nations, from, the, two, regions, Asia-Europe summit, summit begins, begins in, in Beijing, Beijing amid, amid financial, financial crisis, crisis Chinese, Chinese Premier, Premier Wen, Wen Jiabao, Jiabao opened, opened the, the two-day, two-day Asia-Europe, Asia-Europe Meeting, Meeting here, here on, on Friday, Friday with, with the, the summit, summit attended, attended by, by leaders, leaders from, from 43, 43 nations, nations from, from the, the two, two region

Concepts: EthnicGroupOfChinese DenseAsLead Definite_NLAttr
with_UnderspecifiedAgent Parka CityOfCorinthGreece Summit_County_Utah
Crisis Summit_County_Colorado FirstOrderAdministrativeDivision
OpeningAFacility from_UnderspecifiedLocation Region SummitMeeting
isLeaderOf Here AttendingSomething TwoDollarBill_US Friday
EncounteringSomething Ming_NaWen Comedian PremierePerformance Aorta
ChinesePerson CityOfBeijingChina Leader AmsterdamIsland ChineseCuisine
GasClothesDryerConnection meetsWith LesSaintesIslands_Guadeloupe
Summit_County_Ohio ChineseLanguageSet ServingSomeone_HelpingAnAgent
detailedInProp Meeting_SocialGathering Beijing Starting premier isLedBy
Beijing_MunicipalityChina Two_TheTVShow Summit

Running

```

public class Cloud {


    Set<String> directConcepts;
    Set<String> directConceptsCycl;
    Set<String> allConcepts;

    public Cloud(AnnotatedDocument forDoc) {
        directConcepts = new HashSet<String>();

        for (String gram : forDoc.allNGrams()) {
            List<String> concepts = (new CycConceptsForString(gram)).allConcepts();
            if (concepts != null) {
                for (String concept : concepts) {
                    directConcepts.add(concept);
                    CycConcept c = new CycConcept(concept);
                    if (c.hasCycl()) {
                        directConceptsCycl.add(c.getCycl());
                        allConcepts.add(c.getCycl());
                    }
                }
            }
        }

        CycGeneralizations gen1 = new CycGeneralizations(concept);
        if (gen1 != null) {
            List<String> gen1SS = gen1.allConceptsCycl();
            if (gen1SS != null) {
                for (String gen1S : gen1SS) {
                    allConcepts.add(gen1S);
                }
            }
        }
    }
}

```



The Code: Finding the Concepts

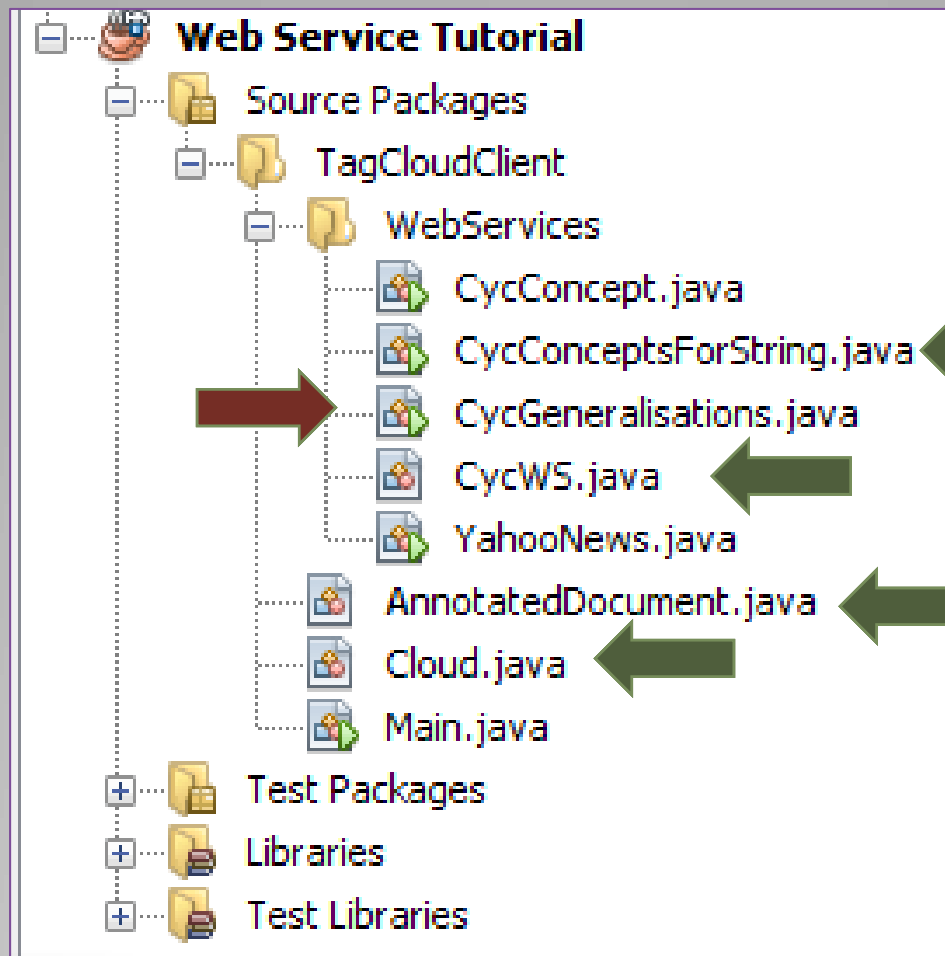
Text: Asia-Europe summit begins in Beijing amid financial crisis. Chinese Premier Wen Jiabao opened the two-day Asia-Europe Meeting here on Friday, with the summit attended by leaders from 43 nations from the two regions.

NGrams: Asia-Europe, summit, begins, in, Beijing, amid, financial, crisis, Chinese, Premier, Wen, Jiabao, opened, the, two-day, Asia-Europe, Meeting, here, on, Friday, with, the, summit, attended, by, leaders, from, 43, nations, from, the, two, regions, Asia-Europe summit, summit begins, begins in, in Beijing, Beijing amid, amid financial, financial crisis, crisis Chinese, Chinese Premier, Premier Wen, Wen Jiabao, Jiabao opened, opened the, the two-day, two-day Asia-Europe, Asia-Europe Meeting, Meeting here, here on, on Friday, Friday with, with the, the summit, summit attended, attended by, by leaders, leaders from, from 43, 43 nations, nations from, from the, the two, two region

Concepts: EthnicGroupOfChinese DenseAsLead Definite_NLAttr
with_UnderspecifiedAgent Parka CityOfCorinthGreece Summit_County_Utah Crisis
Summit_County_Colorado FirstOrderAdministrativeDivision OpeningAFacility
from_UnderspecifiedLocation Region SummitMeeting isLeaderOf Here
AttendingSomething TwoDollarBill_US-Friday EncounteringSomething Ming_NaWen
Comedian-PremierePerformance-Aorta-ChinesePerson CityOfBeijingChina Leader
AmsterdamIsland-ChineseCuisine-GasClothesDryerConnection-meetsWith
LesSaintesIslands_Guadeloupe-Summit_County_Ohio-ChineseLanguageSet
ServingSomeone_HelpingAnAgent-detailedInProp-Meeting_SocialGathering Beijing
Starting premier isLedBy Beijing_MunicipalityChina Two_TheTVShow Summit

Running

Note: Utterly simple tagging; No disambiguation



Where are we?

OpenCyc (Current): [<http://sw.opencyc.org/concept/Mx4rvVjVPZwpEbGdrcN5Y29ycA>]

OpenCyc (Versioned): [<http://sw.opencyc.org/2008/06/10/concept/Mx4rvVjVPZwpEbGdrcN5Y29ycA>]

Search

Mx4rvVjVPZwpEbGdrcN5Y29ycA

OpenCyc Individual: Beijing

Unique ID: [Mx4rvVjVPZwpEbGdrcN5Y29ycA]

English ID:

English Ali:

CityOfBeijingChina, "Peking"]

The capitalCity of China-AtPresent. Note that Peking is the old version of Beijing.

Instance of: capital

Wikipedia:

<http://en.wikipedia.org/wiki/Beijing>

Same as:

<http://dbpedia.org/resource/Beijing>

<http://www.w3.org/2006/03/wn/wn20/instances/synset-Beijing-noun-1>

北京市

A bit of Paper Work: Getting Cycl

```

public CycConcept(String searchExternalId) {
    if (searchExternalId == null) {
        throw new RuntimeException("CycConcept got null input");
    }
}

```

```

URI address = new URI("http", null, "ws.opencyc.org", 80,
    "/webservices/concept/find",
    "str=" + encSearch
    + "&uriType=readable&searchType=externalId", null);
service.addPort(portName, HTTPBinding.HTTP_BINDING, address.toString());
Dispatch<Source> dispatcher = service.createDispatch(portName,
    Source.class, Service.Mode.PAYLOAD);

```

```

service.addPort(portName, HTTPBinding.HTTP_BINDING, address.toString());
Dispatch<Source> dispatcher = service.createDispatch(portName,
    Source.class, Service.Mode.PAYLOAD);
Map<String, Object> requestContext = dispatcher.getRequestContext();
requestContext.put(MessageContext.HTTP_REQUEST_METHOD, new String("GET"));
StreamSource result = (StreamSource) dispatcher.invoke(null);
if (result == null) {
    if (debug) System.out.println(address + " failed to get response for "
        + encSearch + " in CycGeneralisations");
    return;
}

```

```

NodeList resultList = (NodeList) xp.evaluate(
    "/concepts/concept", response, XPathConstants.NODESET);
if (resultList != null && resultList.getLength() > 0) {
    String uri = (String) xp.evaluate("@uri", resultList.item(0), XPathConstants.STRING);
    String conceptID = URLDecoder.decode(uri.replaceFirst(uriPrefix, ""), "UTF-8");
    cycL = conceptID;
}

```

```

static final String uriPrefix = "http://sw.opencyc.org/2008/06/10/concept/en/";

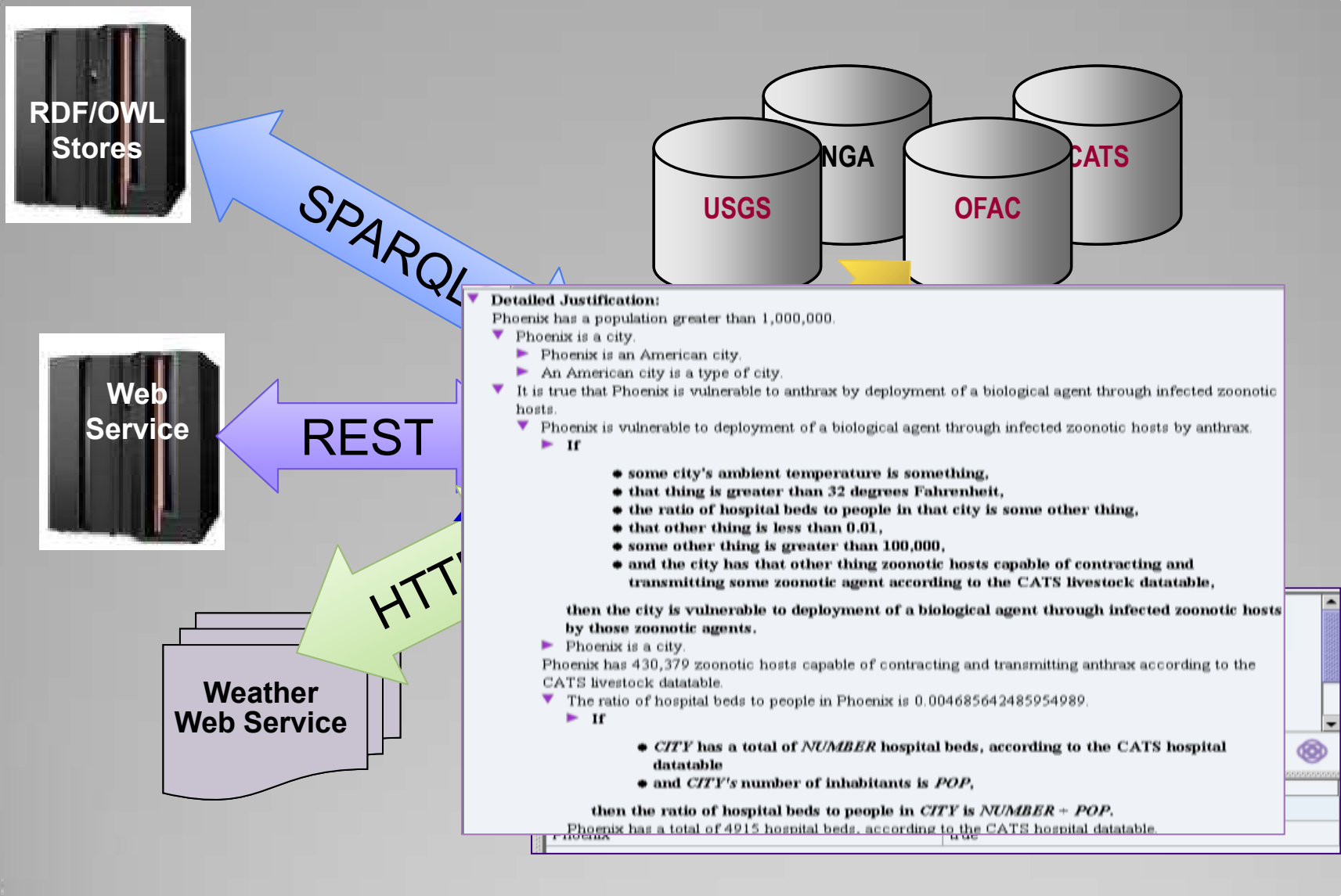
```

Getting Cyc for Concepts



Why do that?

Using Semantic Terms for Data Integration



Find all CABG patients between 1989 and 1998 with an ITA graft that had a prior cardiac catheterization

The screenshot shows the Cyc software interface with a query entered in the 'Queries' tab. The query is: 'Find all CABG patients between 1989 and 1998 with an ITA graft that had a prior cardiac catheterization'. The interface includes a 'Task Info' tab, a 'Document Search' tab, and a 'Concepts' tab. The 'Queries' tab is active, showing the query and a 'Find' button. Below the query, there is a 'Search Results' section with a tree view. The tree view shows the query and its components, including 'Simple Questions' and 'Keyword Search'. The 'Simple Questions' section lists several conditions: '1998 was the date of ____', '1989 was the date of ____', '____ is a patient.', '____ is an internal thoracic artery.', '____ is a coronary bypass.', '____ is a cardiac catheterization.', '____ happened sometime between 1989 and 1998.', '____ happened sometime between 1989 and 1998.', 'The graft ____ uses internal thoracic arteries as a coronary bypass conduit.', '____ was treated during ____.', '____ was put somewhere during ____.', '____ is before ____.', '____ has ____ as a subevent.', and 'The patient ID for ____ is ____.'. The 'Keyword Search' section shows the query: 'Find all operations at CCF between January 1, 1990 and June 30, 1999 where the patient had a prior coronary artery stent placement'. Below the search results, there is a 'CCF Query Library' section with a list of queries: 'Doug Lenat: Generic Medical Research Task Allotment' and 'Search History'. The status bar at the bottom shows 'Status: Idle' and 'Message:'. On the right side of the interface, there is a 'Context: Anytime PSC' section with an 'Edit...' button. Below this, there is a 'Results: All' dropdown and a 'Time limit: 60' dropdown. A 'More...' button is also present. The 'Answers' section shows a table with columns: 'Y 2', 'X 1', 'Y', 'Y 1', 'Thing', 'X 2', and 'X'. The table is currently empty. Below the table, there are buttons for 'Justify', 'Fact Sheet', 'Visualize', and 'Visualize All'. On the far right, there is a vertical toolbar with buttons: 'Ask', 'Save', 'New Tab', and 'Stop'.

Find all CABG patients between 1989 and 1998 with an ITA graft that had a prior cardiac catheterization

ITA

Search Results

Find all CABG patients between 1989 and 1998 with an ITA graft that had a prior cardiac catheterization

Simple Questions

- 1998 was the date of ____.
- 1989 was the date of ____.
- ____ is a patient.
- ____ is an internal thoracic artery.
- ____ is a coronary bypass.
- ____ is a cardiac catheterization.
- ____ happened sometime between 1989 and 1998.
- ____ happened sometime between 1989 and 1998.
- The graft ____ uses internal thoracic arteries as a coronary bypass conduit.
- ____ was treated during ____.
- ____ was put somewhere during ____.
- ____ is before ____.
- ____ has ____ as a subevent.
- The patient ID for ____ is ____.

Keyword Search

Find all operations at CCF between January 1, 1990 and June 30, 1999 where the patient had a prior coronary artery stent placement

CCF Query Library

- Doug Lenat: Generic Medical Research Task Allotment
- Search History

Context: Anytime PSC Edit...

Results: All Time limit: 60 More...

Answers

Y 2	X 1	Y	Y 1	Thing	X 2	X
-----	-----	---	-----	-------	-----	---

Justify Fact Sheet Visualize Visualize All

Status: Idle Message:

Application: Cyc supports access to medical RDF store



OpenCyc Web Services 2:

Generalisation in the Taxonomy

```

public CycGeneralisations(String conceptToFind) {
    if (conceptToFind == null) {throw new RuntimeException("CycConcept got null input"); }

    URI address = new URI("http", null, "ws.opencyc.org", 80,
        "/webservices/concept/taxonomy",
        "externalId=" + encSearch
        + "&relation=typeOf&direction=moreGeneral&conceptDetails=typical", null);
    service.addPort(portName, HTTPBinding.HTTP_BINDING, address.toString());
    Dispatch<Source> dispatcher = service.createDispatch(portName,
        Source.class, Service.Mode.PAYLOAD);

    externalId = encSearch
        + "&relation=typeOf&direction=moreGeneral&conceptDetails=typical", null);
    service.addPort(portName, HTTPBinding.HTTP_BINDING, address.toString());
    Dispatch<Source> dispatcher = service.createDispatch(portName,
        Source.class, Service.Mode.PAYLOAD);
    Map<String, Object> requestContext = dispatcher.getRequestContext();
    requestContext.put(MessageContext.HTTP_REQUEST_METHOD, new String("GET"));
    StreamSource result = (StreamSource) dispatcher.invoke(null);
    if (result==null){
        if (debug) System.out.println(address+" failed to get response for "
            +encSearch+" in CycGeneralisations");
        return;
    }
    InputStream inStream = result.getInputStream();

```

Taxonomy Service: Access

```
InputStream inStream = result.getInputStream();  
if (false) { dumpStream(inStream); } //Debug
```

```
NodeList resultList = (NodeList) xp.evaluate(  
    "/concepts/concept",  
    response, XPathConstants.NODESET);  
if (resultList != null) {  
    int len = resultList.getLength();  
    for (int i = 0; i < len; i++) {  
        String eId = (String) xp.evaluate("externalId", resultList.item(i), XPathConstants.STRING);  
        String cycL = (String) xp.evaluate("cycL", resultList.item(i), XPathConstants.STRING);  
        if (eId != null) {  
            if (conceptGeneralisations == null) {  
                conceptGeneralisations = new ArrayList<String>(defaultConceptCount);  
            }  
            conceptGeneralisations.add(eId);  
        }  
        if (cycL != null) {  
            if (conceptGeneralisationsCycL == null) {  
                conceptGeneralisationsCycL = new ArrayList<String>(defaultConceptCount);  
            }  
            conceptGeneralisationsCycL.add(cycL);  
        }  
        conceptGeneralisationsCycL.add(cycL);  
    }  
}  
}
```

Taxonomy Service Extraction

pub 1



}

pub 1

}

OpenCyc Collection: dog

Unique ID: [Mx4rvVjaoJwpEbGdrcN5Y29ycA]

English ID: [Dog]

English Aliases: ["Canis familiaris", "doggie", "dogs", "hound", "hounds"]

A [BiologicalSpecies](#) (scientific name 'Canis familiaris') that is a specialization of [CanineAnimal](#) (q.v.). Each instance of [Dog](#) is a canine animal that has either been bred to be a domestic pet (see [DomesticatedAnimal](#)) or is a wild canine animal that is not an instance of [Wolf](#), [Fox](#), or any other non-dog specialization of [CanineAnimal](#). Note that although [Dog](#) and [Wolf](#) are considered distinct [BiologicalSpecies](#), instances of the two can and do interbreed successfully. This species classification is therefore unusual, and in some circles, controversial.

A Type of: [Canis](#), [domesticated animal](#)

Instance of: [domesticated animal](#), [some sample kinds of mammal biology topic](#), [species](#)

Subtypes: [bomb sniffing dog](#), [female dog](#), [German Pointer](#), [Herding dog](#), [hound dog](#), [Miscellaneous dog](#), [mixed breed dog](#), [Non-Sporting dog](#), [outside dog](#), [puppy](#), [retriever](#), [setter](#), [Spaniel](#), [sporting dog functional group](#), [standard schnauzer](#), [terrier](#), [Toy dog](#), [Working dog](#)

Instances: [Dash](#)

Wikipedia:

<http://en.wikipedia.org/wiki/Dog>

Same as:

<http://www.w3.org/2006/03/wn/wn20/instances/synset-dog-noun-1>

<http://dbpedia.org/resource/Dog>

<http://umbel.org/umbel/sc/Dog>

Copyright © 2001-2008 Cycorp, Inc.



cA");

Testing Taxonomy Client

Cloud: EthnicGroupOfAsians EthnicGroupOfChinese PhysicalContactEvent
Summit_County_Utah HomoSapiens Emergency NaturalThing SocialGathering
HeadOfState isLeaderOf HeadOfPoliticalParty
ContinuousPhysicalContactSituation Conference Situation-Localized
AsianCitizenOrSubject CountrySubsidiary Event GeographicalPlace
CityOfBeijingChina AmsterdamIsland Situation LandTopographicalFeature
EastAsianCuisine IntelligentAgent meetsWith ChineseLanguageSet
Summit_County_Ohio detailedInProp GeopoliticalEntity Beijing Starting
Meeting_SocialGathering PartiallyTangible isLedBy Beijing_MunicipalityChina
Two_TheTVShow Summit DenseAsLead Definite_NLAttr with_UnderspecifiedAgent
CulturalThing Parka CityOfCorinthGreece Weekday Crisis
Summit_County_Colorado FirstOrderAdministrativeDivision ConvexTangibleObject
OpeningAFacility IntrinsicStateChangeEvent PurposefulAction
from_UnderspecifiedLocation AnimalActivity Region Bill-US EdibleStuff
SummitMeeting GeopoliticalEntityOrRegion Here AttendingSomething
TwoDollarBill_US Friday EncounteringSomething Place HelpingAnAgent
Ming_NaWen Comedian PremierePerformance Aorta ChinesePerson
SurfaceRegion-Underspecified Leader ChineseCuisine GasClothesDryerConnection
HeadOfGovernment EntertainmentPerformance LesSaintesIslands_Guadeloupe
ServingSomeone_HelpingAnAgent premier ManufacturedGoods HumanActivity
Person

Using the generalization



Putting it together

A simple smash-up of Cyc and Yahoo News


```

1 package TagCloudClient;
2 import TagCloudClient.WebServices.YahooNews;
3 /**
4  * @author Michael Witbrock
5  */
6 public class Main {
7     public static void main(String[] args) {
8         try {
9
10             YahooNews n = new YahooNews(args[0]);
11             for (AnnotatedDocument aStory : n.stories()) {
12                 Cloud aCloud = new Cloud(aStory);
13                 System.out.println("INPUT: " + aStory);
14                 System.out.println("\n\nNgrams: " + aStory.allNgrams());
15                 System.out.println("\n\nConcepts: " + aCloud.directConceptsCycLString());
16                 System.out.println("\n\nCloud: " + aCloud);
17             }
18         } catch (Exception e) {
19             System.out.println(e);
20             e.printStackTrace();
21         }
22     }
23 }

```

args[0]=="Europe"

AnnotatedDocument:
Text and NGrams

The Code: Main

```

public YahooNews(String searchTerms, int nStories) {
    stories = new ArrayList<AnnotatedDocument>(defaultStoryCount);
    try {
        String encSearch = URLEncoder.encode(searchTerms, "UTF-8");
        URI nsURI = new URI("urn:yahoo:yn");
        QName serviceName = new QName("YahooSearchService", nsURI.toString());
        QName portName = new QName("YahooSearchPort", nsURI.toString());

        Service service = Service.create(serviceName);

```

```

        URI address = new URI("http", null, "api.search.yahoo.com", 80,
            "/NewsSearchService/V1/newsSearch",
            "appid=" + YahooApplicationID + "&results=" + nStories
            + "&language=en" + "&query=" + encSearch, null);

```

```

        requestContext.put(MessageContext.HTTP_REQUEST_METHOD, new String("GET"));
        StreamSource result = (StreamSource) dispatcher.invoke(null);
        InputStream inStream = result.getInputStream();
        Document response =
            DocumentBuilderFactory.newInstance().newDocumentBuilder().parse(inStream);

        XPathFactory xpf = XPathFactory.newInstance();
        XPath xp = xpf.newXPath();
        NodeList resultList = (NodeList) xp.evaluate(
            "/ResultSet/Result",
            response, XPathConstants.NODESET);
        int len = resultList.getLength();
        for (int i = 0; i < len; i++) {
            String title = (String) xp.evaluate("Title", resultList.item(i),
                XPathConstants.STRING);
            String summary = (String) xp.evaluate("Summary", resultList.item(i),
                XPathConstants.STRING);
            stories.add(new AnnotatedDocument(title + ". " + summary, 2));
        }
    } catch (Exception e) {

```

```

static final String YahooApplicationID =
    "Q.mzkKPV34GJLhdWVgP_ok8d4O4K050uWxYQI7INKcSdIEcW_OEILkYoB6IAeK8-";

```

The Code: Yahoo News

Yahoo News: Asia-Europe summit begins in Beijing amid financial crisis. Chinese Premier Wen Jiabao opened the two-day Asia-Europe Meeting here on Friday, with the summit attended by leaders from 43 nations from the two regions.

NGrams: Asia-Europe, summit, begins, in, Beijing, amid, financial, crisis, Chinese, Premier, Wen, Jiabao, opened, the, two-day, Asia-Europe, Meeting, here, on, Friday, with, the, summit, attended, by, leaders, from, 43, nations, from, the, two, regions, Asia-Europe summit, summit begins, begins in, in Beijing, Beijing amid, amid financial, financial crisis, crisis Chinese, Chinese Premier, Premier Wen, Wen Jiabao, Jiabao opened, opened the, the two-day, two-day Asia-Europe, Asia-Europe Meeting, Meeting here, here on, on Friday, Friday with, with the, the summit, summit attended, attended by, by leaders, leaders from, from 43, 43 nations, nations from, from the, the two, two region

Concepts: EthnicGroupOfChinese DenseAsLead Definite_NLAttr
with_UnderspecifiedAgent Parka CityOfCorinthGreece Summit_County_Utah Crisis
Summit_County_Colorado FirstOrderAdministrativeDivision OpeningAFacility
from_UnderspecifiedLocation Region SummitMeeting isLeaderOf Here
AttendingSomething TwoDollarBill_US-Friday EncounteringSomething Ming_NaWen
Comedian-PremierePerformance-Aorta-ChinesePerson CityOfBeijingChina Leader
AmsterdamIsland-ChineseCuisine-GasClothesDryerConnection-meetsWith
LesSaintesIslands_Guadeloupe-Summit_County_Ohio-ChineseLanguageSet
ServingSomeone_HelpingAnAgent-detailedInProp-Meeting_SocialGathering Beijing
Starting premier isLedBy Beijing_MunicipalityChina Two_TheTVShow Summit

Running

Note: Utterly simple tagging; No disambiguation

INPUT: Miller pens Karlsruhe deal. Karlsruhe have handed goalkeeper Markus Miller a one-year contract extension. The 26-year-old has put pen to paper on the deal that will keep him at the Wildparkstadion until June 2010.

INPUT: Miller pens Karlsruhe deal. Karlsruhe have handed goalkeeper Markus Miller a one-year contract extension. The 26-year-old has put pen to paper on the deal that will keep him at the Wildparkstadion until June 2010.

Ngrams: [Miller, pens, Karlsruhe, deal, Karlsruhe, have, handed, goalkeeper, Markus, Miller, a, one-year, contract, extension, The, 26-year-old, has, put, pen, to, paper, on, the, deal, that, will, keep, him, at, the, Wildparkstadion, until, June, 2010, Miller pens, pens Karlsruhe, Karlsruhe deal, deal Karlsruhe, Karlsruhe have, have handed, handed goalkeeper, goalkeeper Markus, Markus Miller, Miller a, a one-year, one-year contract, contract extension, extension The, The 26-year-old, 26-year-old has, has put, put pen, pen to, to paper, paper on, on the, the deal, deal that, that will, will keep, keep him, him at, at the, the Wildparkstadion, Wildparkstadion until, until June, June 2010]

Concepts: possesses SheetOfPaper DennisMiller AltonGlennMiller DanceBandLeader PuttingSomethingIntoACertainState NewspaperCopy KeepingSomething MaintainingSomething RogerMiller Keeping_ProvidingLodging until SoccerGoalkeeper Miller_County_Missouri MitchMiller Conducting_Administration_Management ContractDocument Backreference_ClassB_NLAttr HumanMaleFirstName NewspaperOrganization HenryMiller_Author Will_County_Illinois ScholarlyArticle Paper June SusanSaintJames PhoneNumberExtension that JonnyLeeMiller Agreement ConformingOnesActionsToSomething subscribers DistributingSomething Definite_NLAttr AcademicComposition ArthurMiller NewspaperSeries HandingSomethingToSomeone Demonstrative_ClassB_NLAttr Pen_TheProgram Handling_Management PenelopeAnnMiller Miller_County_Arkansas PuttingSomethingSomewhere Paper_196 UploadingAFile GeorgeMiller_MovieDirector Excitement Contract ChristaMiller Miller_County_Georgia NewspaperIssue SteveMiller KeepingSomething_RaisingLivingThings MasterP Keeping_ProtectingSomething AnimalPen

Cloud: SheetOfPaper PeriodicalCopy Opaque AltonGlennMiller_DanceBandLeader NewspaperCopy KeepingSomething IntelligentAgentActivity Administration-Management Keeping_ProvidingLodging until SoccerGoalkeeper MediaOrganization CombustibleMaterial MitchMiller Backreference_ClassB_NLAttr Conducting_Administration_Management HumanMaleFirstName CalendarMonth WritingImplement Will_County_Illinois HenryMiller_Author Fibrous Artifact-NonAgentive Paper SoccerPlayer June PhoneNumberExtension StationeryProduct DeviceWithNoMovingParts Policy ConformingOnesActionsToSomething OutdoorLocation subscribers RaisingLivingThings OrganicMaterial ProtectingSomething Definite_NLAttr Management MakingWrittenMaterial ExercisingAuthoritativeControlOverSomething AcademicComposition ArthurMiller IntrinsicStateChangeEvent NewspaperSeries ExpositoryArticle RetainingUserRights IDString ContainerArtifact VisualInformationGeneratingAct CompiledConceptualWork Pen_TheProgram NonfictionalPCW Handling_Management PenelopeAnnMiller Miller_County_Arkansas CWInstantiating Translation-Complete Paper_196 UploadingAFile GeorgeMiller_MovieDirector Excitement PeriodicalIssue Miller_County_Georgia NewspaperIssue SteveMiller AnimalContainer KeepingSomething_RaisingLivingThings NetworkPropagation HumanActivity AnimalPen possesses Foldable Handling-Management DennisMiller PeriodicalSeries PuttingSomethingIntoACertainState MaintainingSomething RogerMiller GregorianMonth Flexible Miller_County_Missouri ProvidingLodging SurfaceRegion-Tangible ContractDocument GuidingAMovingObject FreeSheet NewspaperOrganization LegalAgreement CausingAnotherObjectsTranslationalMotion PropositionalConceptualWork PersonWithOccupation ScholarlyArticle DisposableProduct TransferringPossession Event SusanSaintJames that JonnyLeeMiller Agreement SurfaceRegion-Finite TakingCareOfSomething PublishedConceptualWork DistributingSomething CopyingAComputerFile Translation-LocationChange PurposefulAction HandingSomethingToSomeone Demonstrative_ClassB_NLAttr SurfaceRegion-Bounded PuttingSomethingSomewhere LegalDocument HardcopyInformationBearingObject TemporallyExistingThing Contract ChristaMiller OfficialDocument MasterP PointyEnded Keeping_ProtectingSomething

BUILD SUCCESSFUL (total time: 1 minute 13 seconds)

- Air Force Rome Labs
- ANSER Inc.
- Austin Info Systems
- Claraview, Inc.
- Conceptual Modeling Group
- Daxtron Laboratories, Inc.
- Defense Informa
- Eclectic Systems
- Federal Universi
- Fraunhofer Insti
- Free University
- Galileo Compan
- Gavin Matthews
- Genomic Center
- Georgia State U
- Georgia Tech Co
- Guyren Howe, I
- Harvard Univers
- Hitachi Systems
- Houston VA Med
- Hybrid Systems
- IBM Research
- Image Matters L
- IMISE, Leipzig U
- Institute for Def
- Institute for Lea
- Institute for the Study of Accelerating Change g
- Intelligent Automation, Inc.
- Intelligent Autonomous System Group, University of Munchen
- Internet Technology School

- ISTC-CNR
- Jagiellonian University
- Jennifer Sullivan, Individual
- Jozef Stefan Institute
- Justsystem Corporation
- Karin Kipper Schuler, Individual

Reaching Web 3.0 is a
collaborative, world-wide effort.

- Ohio State University, Computational Linguistics and Language Technology Lab

Research Cyc Licensees

- Ontotext Lab
- Patrick Cassidy, Individual
- Pierluigi Miraglia
- Radboud University
- Raid Limited
- Rensselaer AI and Reasoning Lab

- Richard McCul
- Ricoh Compan
- Robert Kahlert
- Sapio Systems
- Sohar Univers
- SRI
- Stanford Know
- Stanford Logic
- Stanford NLP I
- Stone's Throw
- Stuyvesant Ca
- Sunny Fugate,
- Symphony Lo
- Tel Aviv Unive
- Terra Incognit
- Texas Intern
- Texas Tech Un
- The College of
- TNO-DMV htt
- Tokyo Institut
- Trimtab ConsultingUCLA
- Universidade do Vale do Rio dos Sinos

- Universite Paris
- University degli Studi di Bari
- University of Colorado
- University of Hawaii uofhawaii
- University of Illinois at Urbana-Champaign

Reaching Web 3.0 is a
collaborative and world-wide
effort.

For fun, follow Cyc_ai on
twitter, or play
game.cyc.com.

- Xerox PARC

Research Cyc Licensees

Reaching Web 3.0 is a collaborative effort. Try out LarKC starting next year.

LarKC early adopter release

