Università di Modena e Reggio Emilia
Facoltà di Scienze Fisiche, Informatiche e Matematiche

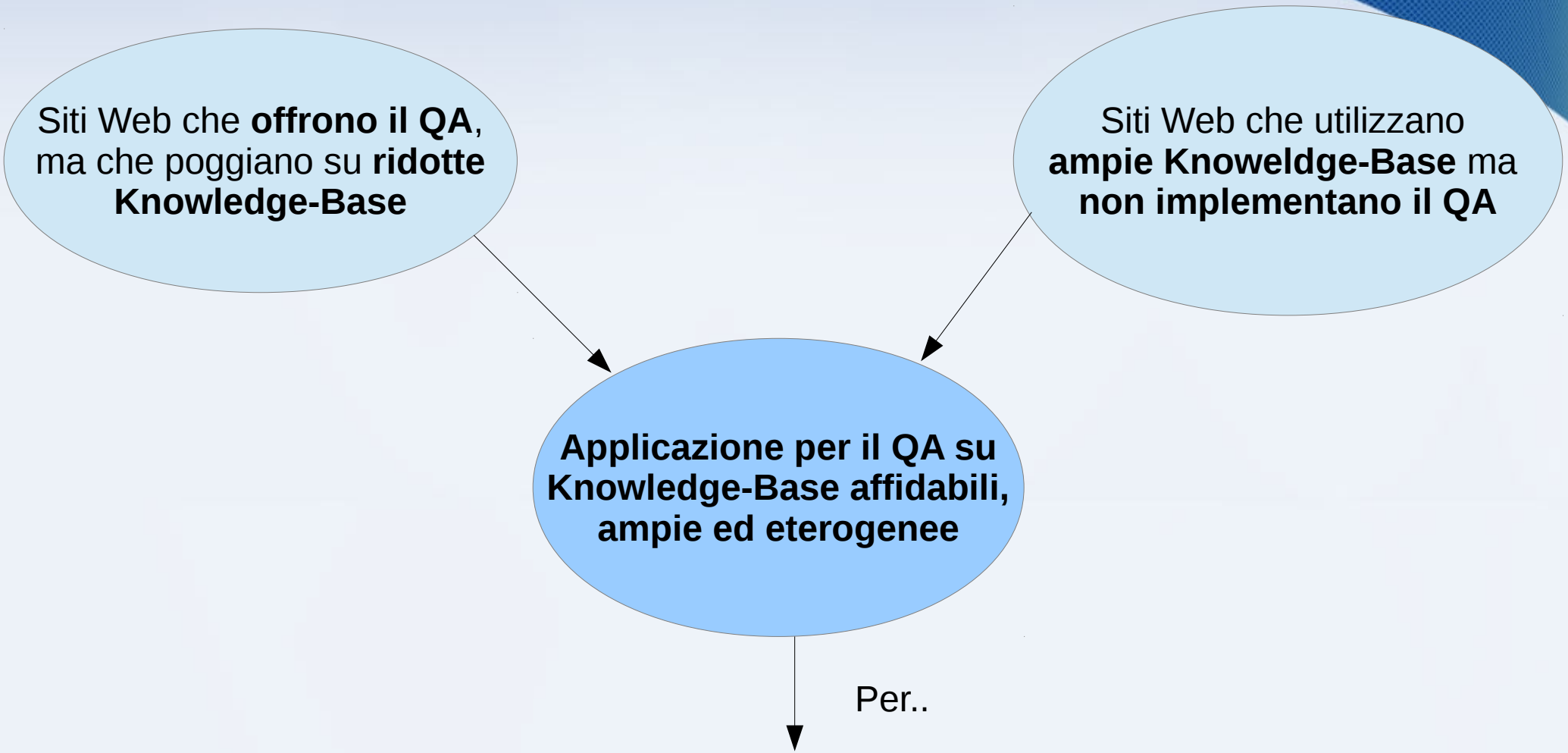
Corso di Laurea in
Informatica

Progetto e sviluppo di un'applicazione di Question Answering su Knowledge Base eterogenee

Tesi di Laurea di
Sonia Moretti

Relatore
Prof. Riccardo Martoglia

Obiettivi della tesi



Consentire all'utente l'**inserimento di domande in NL** e **ottenere direttamente possibili approfondimenti** su siti Web differenti, **riguardanti la risposta** ottenuta.

- **Sorgenti informative utilizzate**
- Progettazione dell'applicazione
- Implementazione
- Risultato
- Conclusioni

	WolframAlpha	CIA	DBpedia	
Modalità di interrogazione:	Natural Language	Nomi Stati mondiali	Parole chiave	Parole chiave
Lingua di interrogazione:	Inglese	Inglese	Inglese	Inglese
Cosa offre:	Risposte di carattere scientifico riguardanti tutte le categorie di sapere	Informazioni geografiche, politiche ed economiche riguardanti gli Stati mondiali	Informazioni riguardanti tutte le categorie di sapere	Informazioni riguardanti diverse categorie di sapere
KB scaricabile:	No	Si, tramite la pagina HTML	Si, in formato RDF	Si, in formato RDF

WolframAlpha

- QA disponibile per diversi argomenti
- Lingua: inglese
- Knowledge base non scaricabile

- Informazioni Stati mondiali
- Lingua: inglese
- Knowledge-Base scaricabile

- Informazioni riguardanti molteplici categorie di sapere
- Lingua: inglese
- Knowledge-Base scaricabile in RDF

About: Canberra

An Entity of Type : municipality, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Canberra is the capital city of Australia. With a population of 367,000, it is Australia's largest inland city and the eighth-largest city overall. The city is located at the northern end of the Australian Capital Territory (ACT), 280 km south-west of Sydney, and 660 km north-east of Melbourne. A resident of Canberra is known as a "Canberran".

Property	Value
dbpedia-owl:abstract	▪ Canberra är Australiens huvudstad. Den är belägen på Staden har drygt 320 000 invånare (2005) och är Austr ▪ Canberra (también admitido en español Camberra; del la capital de Australia, con una población superior a lo Australiana, ubicada en la parte norte del mismo, 300 noreste de Melbourne. La situación de Canberra fue se como un compromiso entre Sídney y Melbourne, las do australianas, ya que fue construida desde cero, como para el diseño de la ciudad, se seleccionó el proyecto Chicago y la construcción comenzó en 1913. El diseño jardín" e incorpora en diferentes áreas vegetación nat "capital del arbusto". A pesar de que el crecimiento y guerras mundiales y la Gran Depresión, a partir de la S próspera. En tanto que sede del Gobierno de Australia Corte Suprema de Australia, además de numerosas ag instituciones sociales y culturales de importancia naci al PBI y aporta una amplia oferta laboral. Canberra es para los extranjeros. ▪ Camberra (em inglês escrito Canberra, que em dialeto conhecido") é a capital da Austrália. Com uma populaç australiano e a oitava maior do país. O local para a co compromisso para acabar com a rivalidade de Sydney um concurso internacional de design da cidade e o pro Griffin foi o escolhido e a construção começou em 191
dbpedia-owl:country	▪ dbpedia:Australia
dbpedia-owl:populationTotal	▪ 367752 (xsd:integer)
dbpedia-owl:thumbnail	▪ http://upload.wikimedia.org/wikipedia/commons/thumb/
dbpedia-owl:wikiPageExternalLink	▪ http://books.google.com/books?id=xSv2HvrNSo0C ▪ http://www.gim.act.gov.au/website/actlocate/Viewer.htm ▪ http://www.wikimapia.org/#y=-35310000&x=14913000 ▪ http://www.actmapi.act.gov.au/ ▪ http://www.act.gov.au/ ▪ http://www.canberra100.com.au/ ▪ http://www.visitcanberra.com.au/ ▪ http://www.australia.com/explore/cities/canberra.aspx

Freebase

Freebase | Find... | Browse | Query | Help | Sign In or Sign Up | English

Canberra en

Created by chaitanyakuber on 7/18/2007

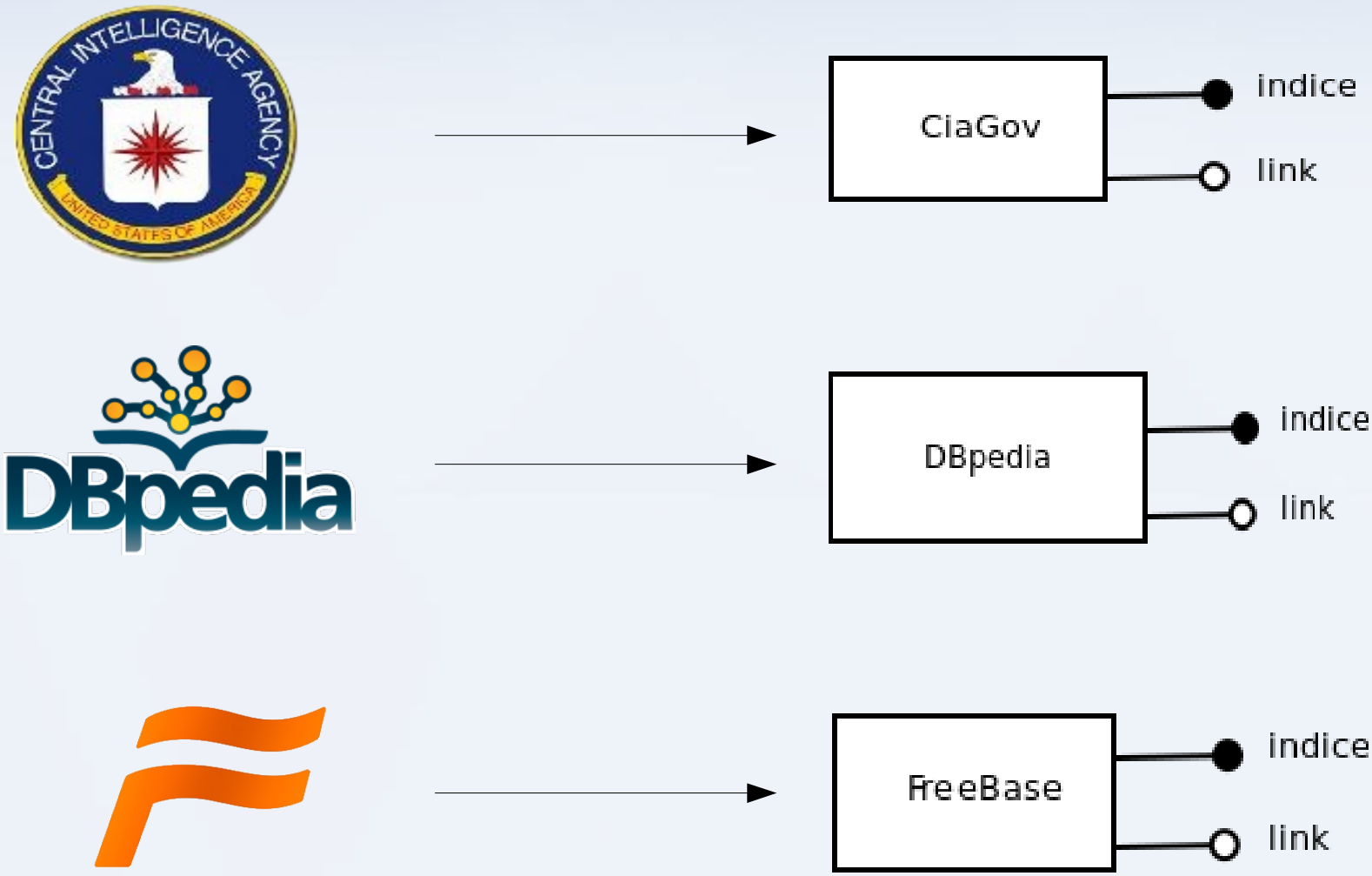
mid: /m/0dp90 **notable type:** /location/citytown **notable for:** /location/citytown **on the web:** wikipedia.org

Canberra is the capital city of Australia. With a population of 374,245, it is Australia's largest inland city and the eighth-largest city overall. The city is located at the northern end of the Australian Capital Territory, 280 km south-west of Sydney, and 660 km north-east of Melbourne. A resident of Canberra is known as a "Canberran". The site of Canberra was selected for the location of the nation's capital in 1908 as a compromise between rivals Sydney and Melbourne, Australia's two largest cities. It is unusual among Australian cities, being an entirely planned city outside of any state, similar to the American Federal District of Columbia. Following an international contest for the city's design, a blueprint by the Chicago architects Walter Burley Griffin and Marion Mahony Griffin was selected and construction commenced in 1913. The Griffins' plan featured geometric motifs such as circles, hexagons and triangles, and was centred on axes aligned with significant topographical landmarks in the Australian Capital Territory. The city's design was influenced by the garden city movement and incorporates significant areas of natural vegetation that have earned Canberra the title of the "bush capital". The growth and development of Canberra were hindered by the World Wars and the Great Depression, which exacerbated a series of planning disputes and the ineffectiveness of a procession of bodies that were created in turn to oversee the development of the city. The national capital emerged as a thriving city after World War II, as Prime Minister Robert Menzies championed its development and the National Capital Development Commission was formed with executive powers. Although the Australian Capital Territory is now self-governing, the federal government retains some influence through the National Capital Authority. Wikipedia [-]

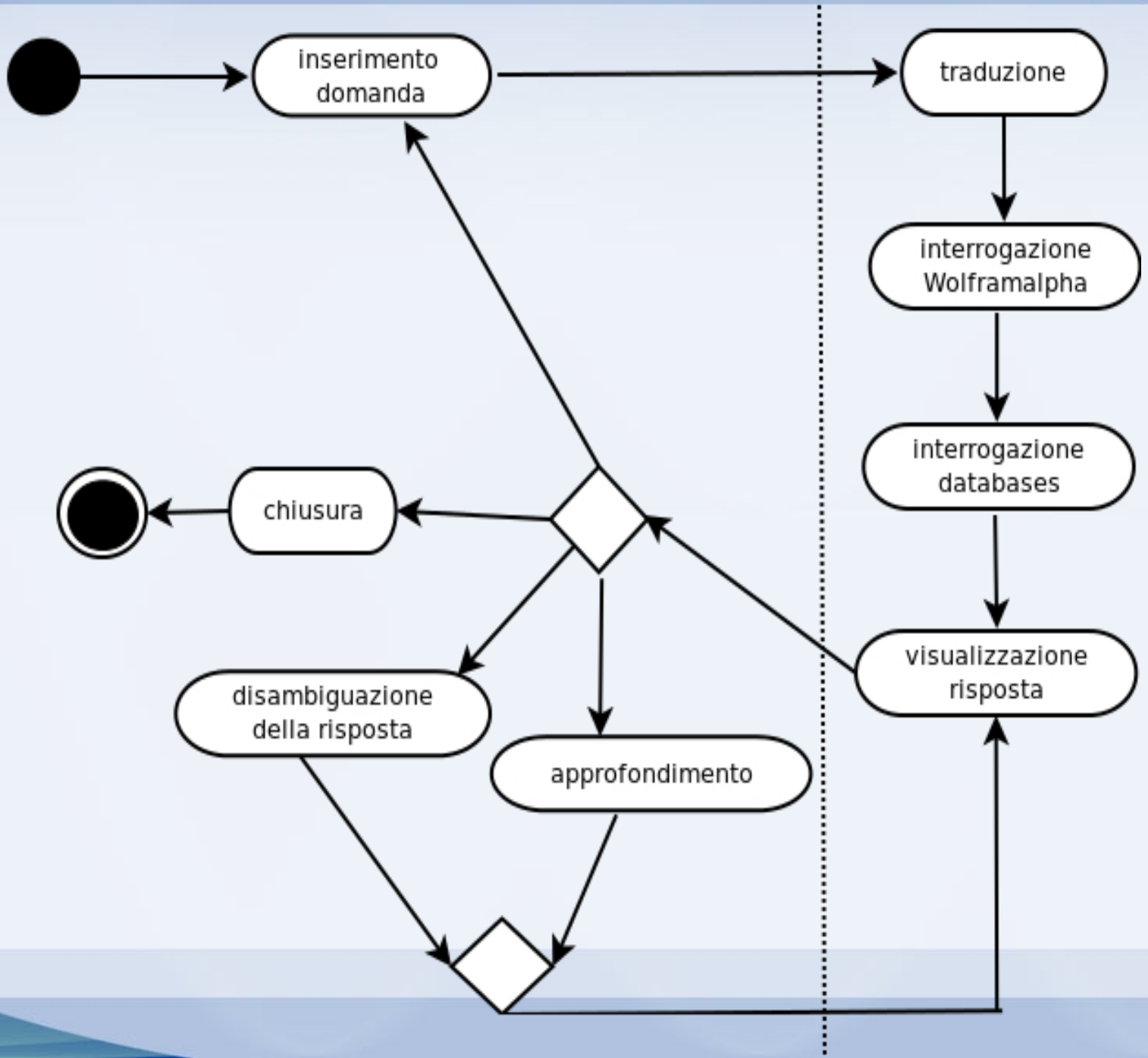
- Informazioni riguardanti diversi categorie di sapere, risultati migliori riguardanti persone e personaggi
- Lingua: inglese
- Knowledge-Base scaricabile in RDF

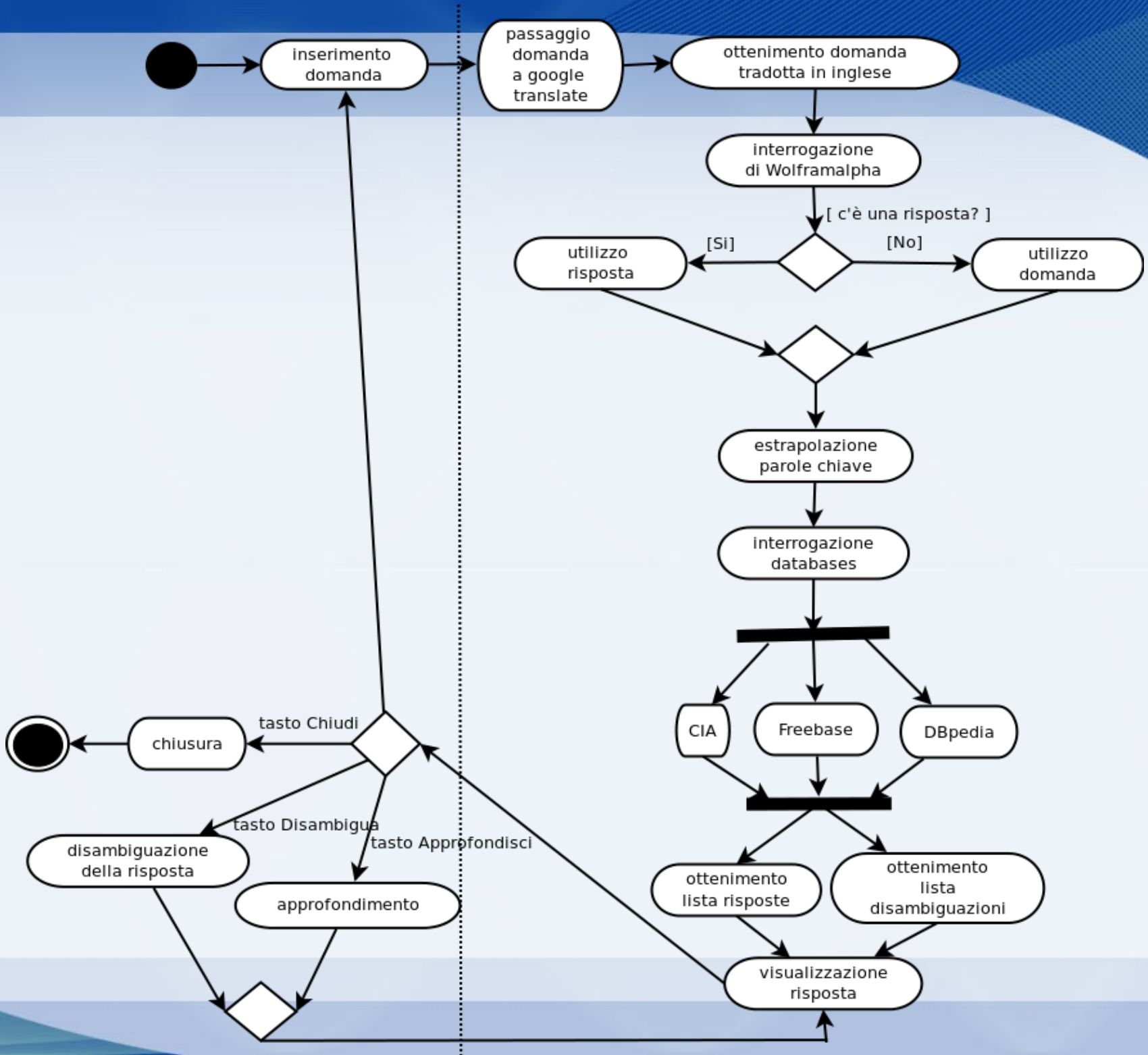
- Sorgenti informative utilizzate
- **Progettazione dell'applicazione**
- Implementazione
- Risultato
- Conclusioni

I database


CIAGov
indice
link
DBpedia
indice
link
FreeBase
indice
link

Activity Diagram




inserimento domanda
passaggio domanda a google translate
ottenimento domanda tradotta in inglese
interrogazione di Wolframalpha
[c'è una risposta?]
[Si]
[No]
utilizzo risposta
utilizzo domanda
estrapolazione parole chiave
interrogazione databases
CIA
Freebase
DBpedia
ottenimento lista risposte
ottenimento lista disambiguazioni
visualizzazione risposta
tasto Chiudi
chiusura
tasto Disambigua
tasto Approfondisci
disambiguazione della risposta
approfondimento

- Sorgenti informative utilizzate
- Progettazione dell'applicazione
- **Implementazione**

  → **Cenni sulla fase di sviluppo**
- Risultato
- Conclusioni

Creazione dei database

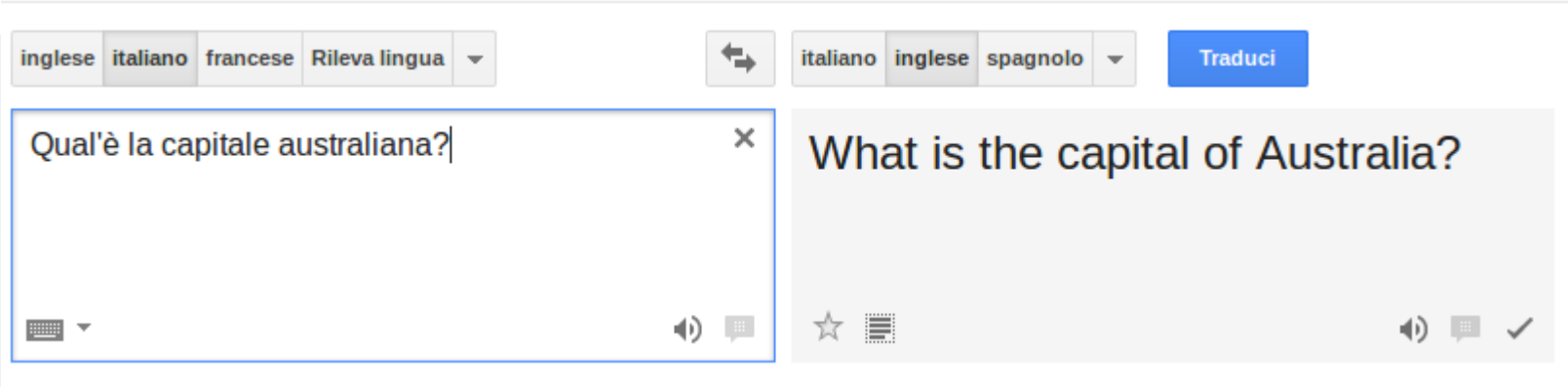
A partire dai dataset scaricati in **RDF**

“modello standard per lo scambio di dati sul Web”

Alla creazione dei database con **SQLite** in Python

“SQLite: libreria che implementa un leggero database disk-based che non necessita di un processo server separato.”

Traduzione della domanda in NL



Utilizzo **URL Request** e analizzo la **pagina HTML** risultatami

NLTK – Natural Language Toolkit

'What is the capital of Australia?'

→ **tokenizzazione**:
['What', 'is', 'the', 'capital', 'of', 'Australia', '?']

→ **rimozione stopwords e stemming:**
['capital', 'Australia']

Algoritmi

- Scelta del **risultato** da mostrare per ogni sito Web→ Algoritmo che utilizza criteri di **maggior numero di parole in comune** e **minor numero di parole totali**

es. 'Canberra', 'Australia'

  - CIA: 'Australia' →...the-world-factbook/geos/as.html
  - DBpedia: 'Canberra', 'Australia' → .../resource/Canberra,_Australia

Algoritmi

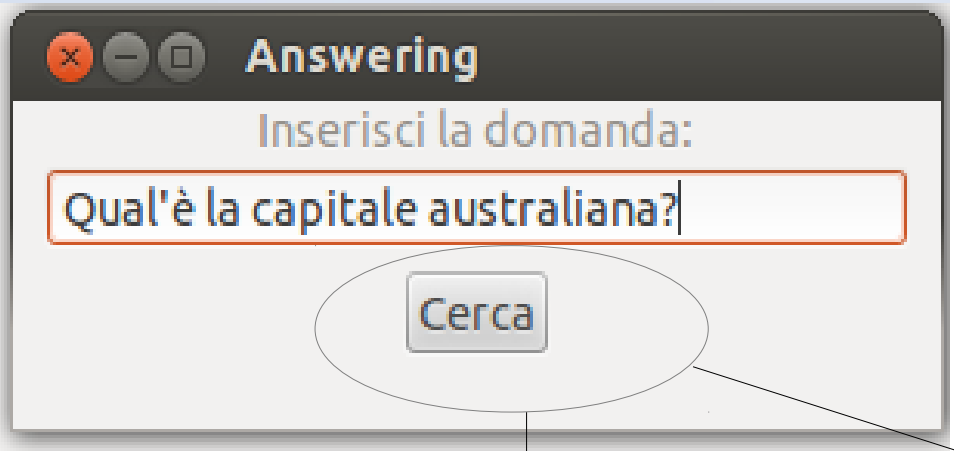
- Scelta delle **disambiguazioni** per una data entità → Algoritmo che utilizza criteri di **'forma' di disambiguazione**: ' ( ) '

  es. 'What is a mouse?' → 'mouse'

  - DBpedia: 'Mouse (computing)', 'Mouse (rodent)', 'Mouse (anime)', etc.
  - Freebase: 'Mouse (set theory)', 'Mouse (manga)', etc.

- Sorgenti informative utilizzate
- Progettazione dell'applicazione
- Implementazione
- **Risultato**
- Conclusioni

1° esempio


Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca

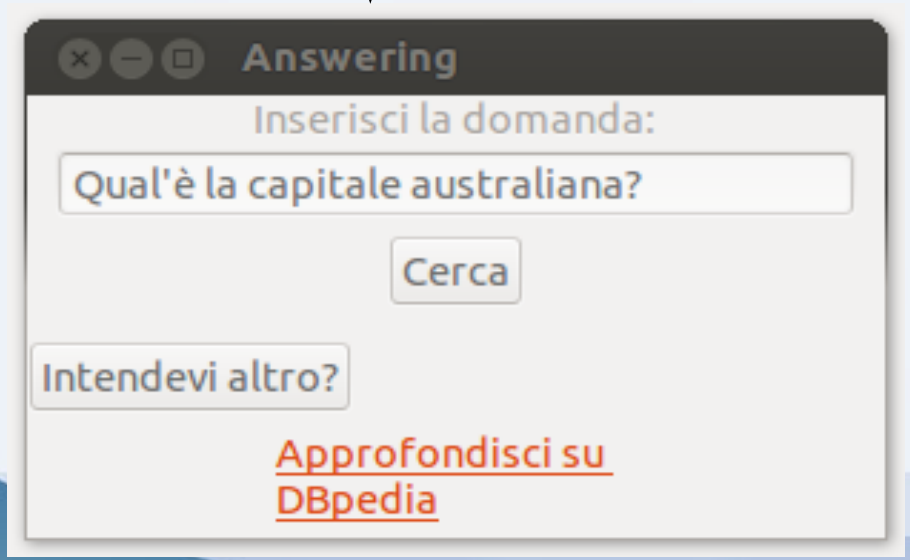
Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca
Intendevi altro?
Approfondisci su CiaGov

WolframAlpha computational knowledge engine
what is the capital of Australia?
Examples
Random
Assuming "Australia" is a country | Use the input as a word instead
Input interpretation:
Australia
capital city
Result:
Canberra, Australian Capital Territory, Australia
Population:
Show history
city population
351 868 people (country rank: 8th) (2009 estimate)
Location:
World map
Show coordinates

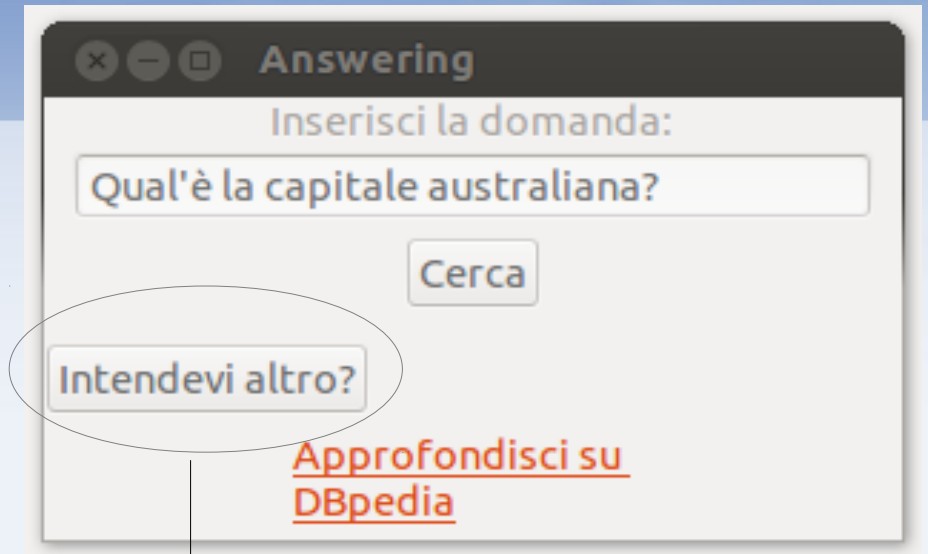
...

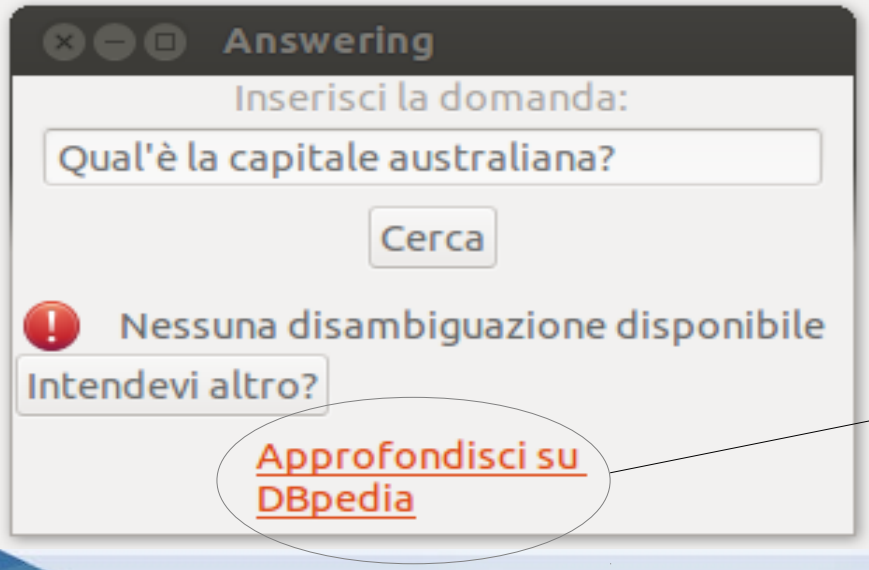
Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca
Intendevi altro?
Approfondisci su CiaGov

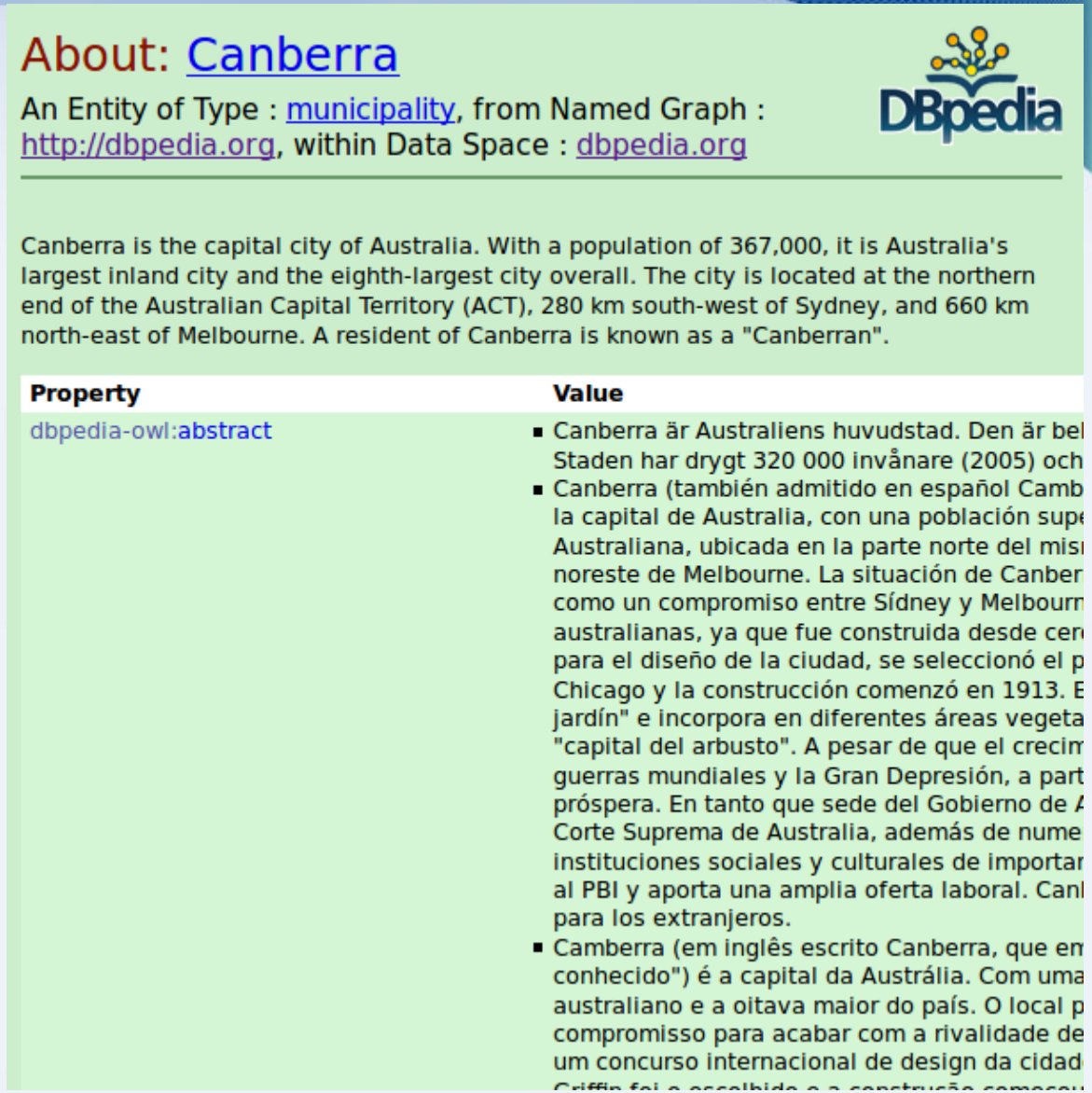
THE WORLD FACTBOOK
Please select a country to view
ABOUT
REFERENCES
APPENDICES
FAQs
CONTACT
VIEW TEXT/LOW BANDWIDTH VERSION
DOWNLOAD PUBLICATION
AUSTRALIA-OCEANIA :: AUSTRALIA
PAGE LAST UPDATED ON MARCH 11, 2014
VIEW 64 PHOTOS OF AUSTRALIA
INDONESIA
PAPUA NEW GUINEA
TIMOR-LESTE
INDIAN OCEAN
Darwin
Cairns
Townsville
Port Hedland
Dampier
Alice Springs
Mackay
Gladstone
Brisbane
Perth
Fremantle
Newcastle
Sydney
Adelaide
CANBERRA
Melbourne
Launceston
Tasmania
Hobart
EXPAND ALL
Introduction :: AUSTRALIA
Geography :: AUSTRALIA
People and Society :: AUSTRALIA
Government :: AUSTRALIA
Economy :: AUSTRALIA
Energy :: AUSTRALIA
Communications :: AUSTRALIA
Transportation :: AUSTRALIA
Military :: AUSTRALIA
Transnational Issues :: AUSTRALIA
EXPAND ALL


Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca
Intendevi altro?
Approfondisci su DBpedia

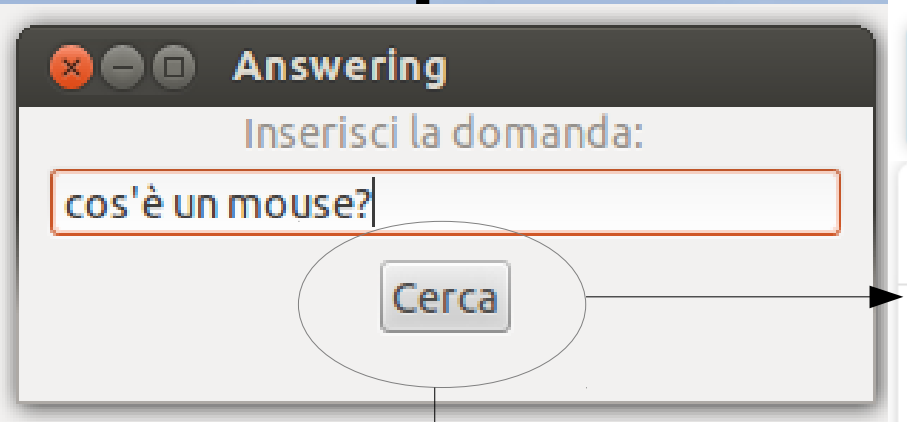
...

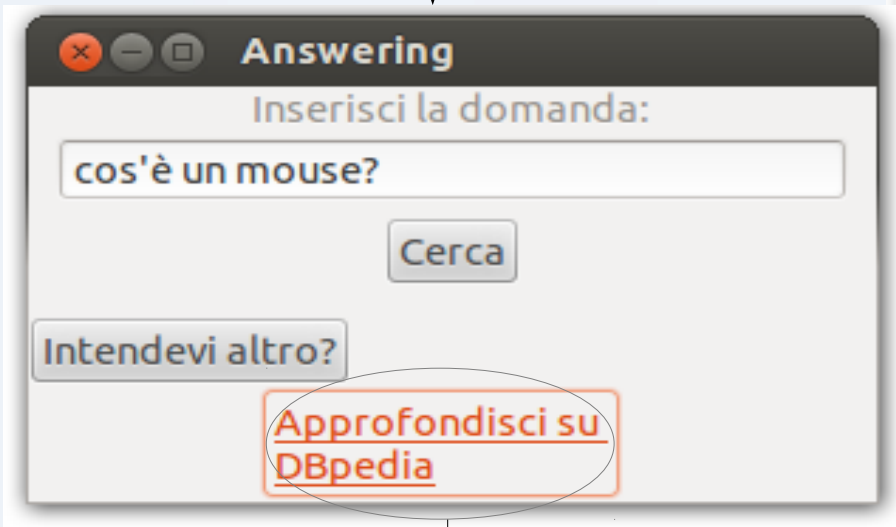

Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca
Intendevi altro?
Approfondisci su DBpedia


Answering
Inserisci la domanda:
Qual'è la capitale australiana?
Cerca
Nessuna disambiguazione disponibile
Intendevi altro?
Approfondisci su DBpedia

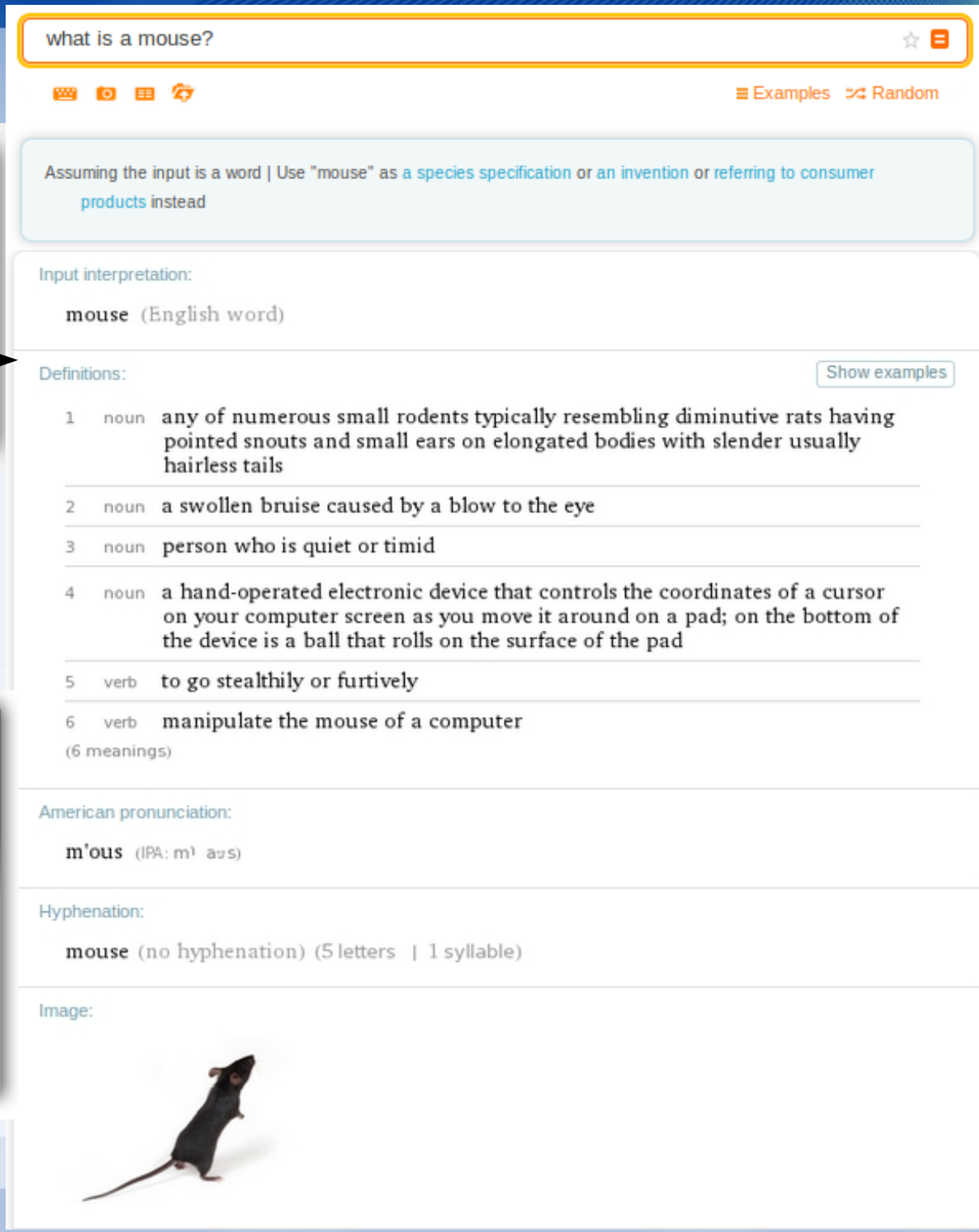

About: Canberra
An Entity of Type : municipality, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org
DBpedia
Canberra is the capital city of Australia. With a population of 367,000, it is Australia's largest inland city and the eighth-largest city overall. The city is located at the northern end of the Australian Capital Territory (ACT), 280 km south-west of Sydney, and 660 km north-east of Melbourne. A resident of Canberra is known as a "Canberran".
Property
Value
dbpedia-owl:abstract
Canberra är Australiens huvudstad. Den är bel
Staden har drygt 320 000 invånare (2005) och
Canberra (también admitido en español Camb
la capital de Australia, con una población supe
Australiana, ubicada en la parte norte del misı
noreste de Melbourne. La situación de Canber
como un compromiso entre Sídney y Melbourn
australianas, ya que fue construida desde cer
para el diseño de la ciudad, se seleccionó el p
Chicago y la construcción comenzó en 1913. E
jardín" e incorpora en diferentes áreas vegeta
"capital del arbusto". A pesar de que el crecim
guerras mundiales y la Gran Depresión, a part
próspera. En tanto que sede del Gobierno de A
Corte Suprema de Australia, además de nume
instituciones sociales y culturales de importar
al PBI y aporta una amplia oferta laboral. Canl
para los extranjeros.
Camberra (em inglês escrito Canberra, que em
conhecido") é a capital da Austrália. Com uma
australiano e a oitava maior do país. O local p
compromisso para acabar com a rivalidade de
um concurso internacional de design da cidad

2° esempio


Answering
Inserisci la domanda:
cos'è un mouse?
Cerca


Answering
Inserisci la domanda:
cos'è un mouse?
Cerca
Intendevi altro?
Approfondisci su DBpedia

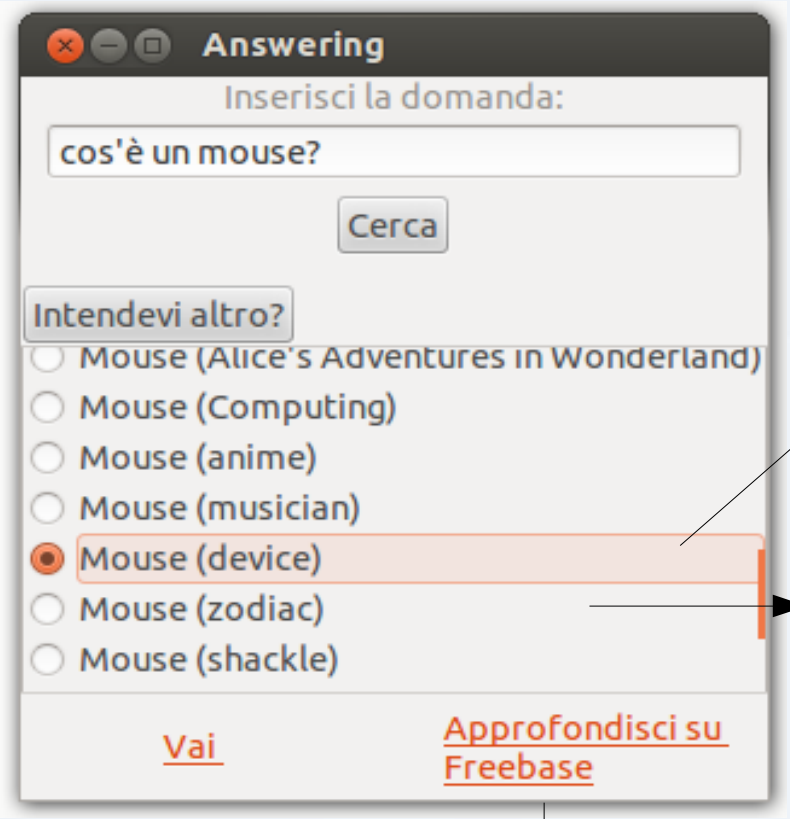
...


what is a mouse?
Examples
Random
Assuming the input is a word | Use "mouse" as a species specification or an invention or referring to consumer products instead
Input interpretation:
mouse (English word)
Definitions:
Show examples
1 noun any of numerous small rodents typically resembling diminutive rats having pointed snouts and small ears on elongated bodies with slender usually hairless tails
2 noun a swollen bruise caused by a blow to the eye
3 noun person who is quiet or timid
4 noun a hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad; on the bottom of the device is a ball that rolls on the surface of the pad
5 verb to go stealthily or furtively
6 verb manipulate the mouse of a computer
(6 meanings)
American pronunciation:
m'ous (IPA: mˈaʊs)
Hyphenation:
mouse (no hyphenation) (5 letters | 1 syllable)
Image:

About: Mus (genere)

An Entity of Type : species, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Mus è un genere di Roditori della famiglia dei Muridae.



About: Mouse

An Entity of Type : PhysicalEntity100001930, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

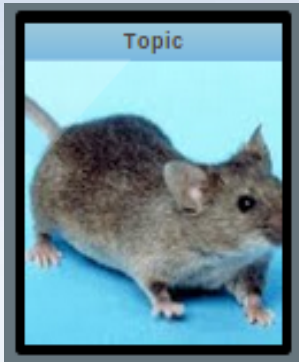
Il mouse è un dispositivo in grado di inviare un input ad un computer in modo tale che ad un suo movimento ne corrisponda uno analogo di un indicatore sullo schermo detto puntatore. È inoltre dotato di uno o più tasti ai quali possono essere assegnate varie funzioni. Esistono due possibili spiegazioni riguardo l'etimologia del nome. La più comune è la parola mouse in relazione alla somiglianza del dispositivo con il roditore.

About: Topo (zodiaco cinese)

An Entity of Type : Animal100015388, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Il Topo era ritenuto, nei tempi antichi, protettore e portatore di prosperità materiale. È un animale associato con l'aggressione, la ricchezza, il fascino e l'ordine, ma anche con la morte, la guerra, l'occulto, la peste e le atrocità.

...

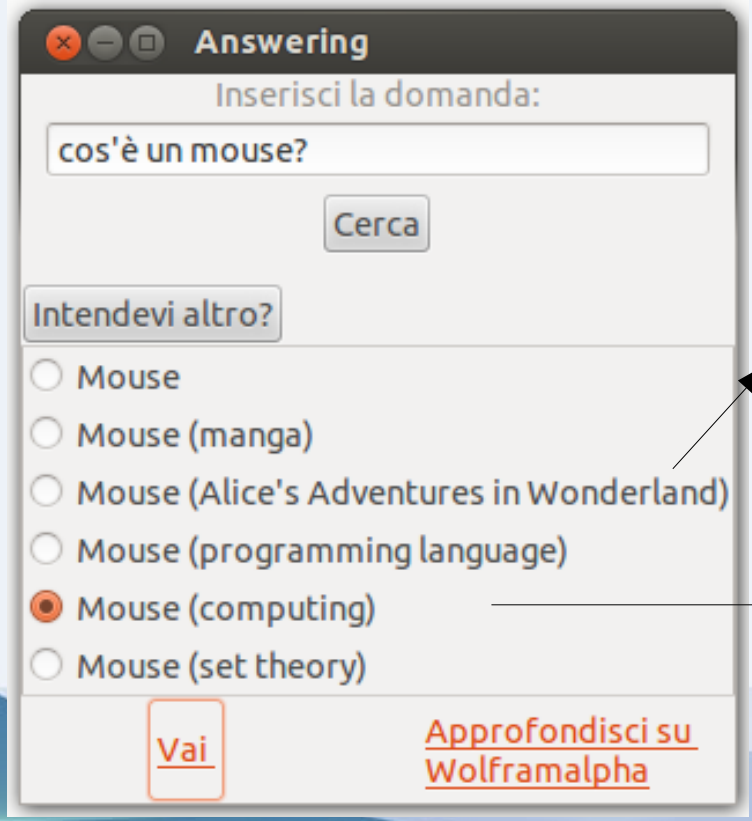


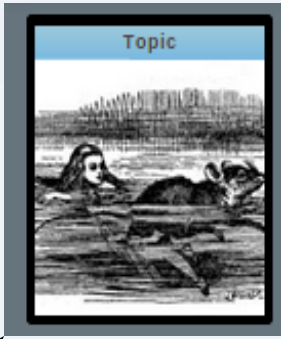
Mouse en

Created by doconnor on 3/10/2010

mid: /m/04rmv **notable type:** /biology/animal **notable for:** /biology/animal **on the web:** wikipedia.org

A mouse is a small mammal belonging to the order of rodents, characteristically having a pointed snout, small rounded ears, and a long naked or almost hairless tail. The best known mouse species is the common house mouse. It is also a popular pet. In some places, certain kinds of field mice are also common. They are known to invade homes for food and occasionally shelter. The American white-footed mouse and the deer mouse, as well as other common species of mouse-like rodents around the world, also sometimes live in houses. These, however, are in other genera. Cats, wild dogs, foxes, birds of prey, snakes and even certain kinds of arthropods have been known to prey heavily upon mice. Nevertheless, because of its remarkable adaptability to almost any environment, the mouse is one of the most successful mammalian genera living on Earth today. Mice can at times be vermin, damaging and eating crops, causing structural damage and spreading diseases through their parasites and feces. In North America, breathing dust that has come in contact with mouse excrement has been linked to hantavirus, which may lead to Hantavirus Pulmonary Syndrome. Wikipedia [-]





Mouse en

mid: /m/02rfcq1 **notable type:** /book/book_character **notable for:** /book/book_character

The Mouse is a fictional character in Alice's Adventures in Wonderland by Lewis Carroll. He ap
Chapter III "A Caucus-Race and a Long Tale". Alice, the eponymous heroine in the book, first t
pool of her own tears, having shrunk in size: `O Mouse, do you know the way out of this pool?
Mouse!' Wikipedia [-]

Mouse en

mid: /m/020lf **notable type:** /computer/computer_peripheral_class **notabl**
wikipedia.org

In computing, a mouse is an input device that detects two-dimensional motion
the motion of a pointer on a display, which allows for fine control of a graphical
held in one's hand, with one or more buttons. Mice often also feature other ele
additional control and dimensional input. Wikipedia [-]

- Sorgenti informative utilizzate
- Progettazione dell'applicazione
- Implementazione
- Risultato
- **Conclusioni**

Conclusioni

È stata realizzata un'applicazione che consente l'inserimento di **domande in NL** in **diverse lingue** e che sfrutta il **QA** per ottenere validi risultati interrogando anche **Knowledge-Base eterogenee**, offrendo inoltre la possibilità di **disambiguare il risultato** ottenuto.

Sviluppi futuri

Possibili ampliamenti:

- Numero di Knowledge-Base utilizzate
- Algoritmi specifici per il QA
- Migliorare la traduzione nella lingua inglese

GRAZIE PER L'ATTENZIONE