

Course: Common Sense Reasoning

8. Collective Acquisition of Common Sense Knowledge

Martin Molina



Common sense knowledge can be acquired collectively

- Acquiring common sense knowledge manually requires a considerable effort
- It is possible to facilitate this task with the participation of thousands of volunteers
- There are successful experiences to acquire other type of content with collaborative work (e.g. Wikipedia)
- Three representative cases are presented:
 - Openmind project
 - Verbosity
 - Wikidata

Openmind is a common sense knowledge base created by thousands of volunteers

- The project started in 1999 and was active until 2013
- The goal of the project was collecting common sense knowledge from volunteers on the internet
- One of the project founders at the MIT was Push Singh



Push Singh

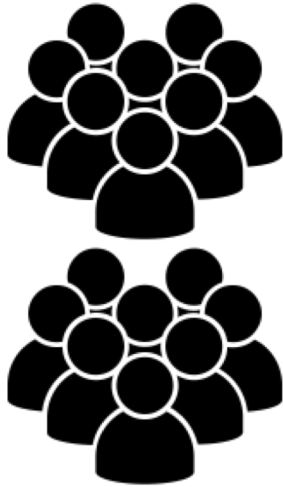
[Singh, 2002]

[Singh et al., 2002]

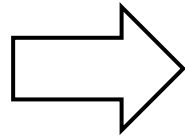
[Singh et al., 2004]

Thousands of volunteers collaborated in the development of Openmind

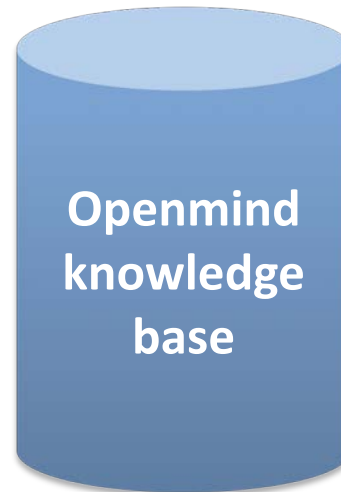
Volunteers



Over 15.000 participants



Collaborative work



Over a 1 million sentences

Openmind used a web site with natural language templates

- A web site gathered descriptions by allowing participants to fill in natural language templates:

_____ is for _____

Your make _____ by _____

You are likely to find _____ in _____

- For example:

A guitar is for making music

Your make an apple pie by baking it

You are likely to find spinach in a supermarket

Similar concepts

[fun](#), [music](#), [play](#), [entertainment](#), [drinking](#), [making music](#), [playing music](#), [singing](#), [entertaining someone](#), [food](#)

Current knowledge

→ An activity singers can do is create music	by chan	Score: 5		
→ The effect of playing the violin is to create music	by sshearer	Score: 4		
→ A wind instrument is used to create music	by butina	Score: 3		
→ a harmonica is for creating music	by openmind	Score: 2		
→ a flute is used for creating music	by ghanney	Score: 2		
→ creating music is for pleasure	by dsmith24	Score: 2		
→ percussion instruments is used for creating music .	by kwitzel	Score: 2		
→ playing a harp is for creating music	by openmind	Score: 2		

Page 1 of 7 | [Next](#) | [Last](#) (51 total)

Open Mind wants to know...

<input type="text" value="creating music"/>	is {%}similar to	<input type="text" value="music"/>	<input type="button" value="+"/>	<input type="button" value="-"/>
<input type="text" value="music"/>	is {%}similar to	<input type="text" value="creating music"/>	<input type="button" value="+"/>	<input type="button" value="-"/>
<input type="text" value="creating music"/>	are	<input type="text" value="relaxing"/>	<input type="button" value="+"/>	<input type="button" value="-"/>

Places to start

Concepts
[go to the grocery store](#), [become smarter](#), [void](#), [seeing friends](#), [boiling water](#), [practice equipment](#), [insulation](#), [opinions](#), [a burden](#)

Vote on these statements...

- a [cat](#) can [kill birds](#)
- [having a conversation](#) is for [Communicating](#)
- Kinds of [religions](#) : [buddhism](#)
- [Some mushrooms](#) are [poisonous](#)
- a [blanket](#) can be used for [sleeping](#)
- You are likely to find [apples](#) in [a house](#).
- [smoking cigarettes](#) is [bad for you](#)
- [Some religions](#) have [ministers](#)
- A [boat](#) is [a form of transportation](#)
- [doing work](#) requires [energy](#)

Feedback

Add new knowledge

MadeOf

What is it made of?

IsA

What kind of thing is it?

UsedFor

What do you use it for?

CapableOf

What can it do?

PartOf

What is it part of?

DefinedAs

How do you define it?

CreatedBy

How do you bring it into existence?

HasFirstSubevent

What do you do first to accomplish it?

HasLastSubevent

What do you do last to accomplish it?

HasPrerequisite

What do you need to do first?

AtLocation

Where would you find it?

MotivatedByGoal

Why would you do it?

Desires

What does it want?

CausesDesire

What does it make you want to do?

Causes

What does it make happen?

HasSubevent

What do you do to accomplish it?

HasProperty

What properties does it have?

ReceivesAction

What can you do to it?

Add new knowledge

Example statements

- [bottles](#) are often made of [plastic](#) 👍👍
- [a book](#) can be made of [paper](#) 👍👍
- [Bottles](#) are usually made of [glass](#) 👍👍
- [Windows](#) are usually made of [glass](#) 👍👍
- [cheese](#) is made from [milk](#) 👍👍

Teach OpenMind another statement of this type.

MadeOf: What is it made of?

<input type="text"/>	can be made of	<input type="text"/>	Teach OpenMind
<input type="text"/>	may be made of	<input type="text"/>	Teach OpenMind
<input type="text"/>	are made of	<input type="text"/>	Teach OpenMind
<input type="text"/>	is made of	<input type="text"/>	Teach OpenMind
<input type="text"/>	can be made from	<input type="text"/>	Teach OpenMind

Ad-hoc categories

Enter a few words in the box below, separated by commas.

Here are concepts and features that are similar to this category:

Similar concepts	
a person	53031095.33
a man	7488801.52
a human	4511502.63
something	4009097.76
a dog	1445020.30
children	1230971.13
a cat	726000.31
A woman	614467.42
a computer	611903.60
animals	479845.15

Similar features	
__ is similar to person.	3145755.43
A person is similar to __.	3077592.87
You are likely to find __ in a hospital.	2341182.58
__ can love	2282036.37
__ can walk the dog	2207500.92
You are likely to find __ in a shop.	2157138.26
__ can water plants	2129660.98
You are likely to find __ in a zoo.	2098355.20
You are likely to find __ in your desk.	2036646.25
You are likely to find __ in a museum.	1991164.81

What are the strengths and the weaknesses of Openmind?

- Strengths
 - Contains over a million sentences corresponding to universals and implicit common sense knowledge
 - A manual evaluation showed an acceptable average quality of 4.12 (from 1 to 5)
 - 84% of the statements are items that most people know
- Weaknesses
 - Limited expressivity of the representation (based on NL templates)
 - Facts that have the same meaning are represented in a different way
 - The knowledge base is incomplete and heterogeneous

Computer games may be used to acquire from volunteers common sense knowledge

- Examples of computer games for knowledge acquisition:
 - Verbosity
 - Gecka


Verbosity is a game of Games with a Purpose (gwap.com) designed to collect common sense knowledge

gwap ESP Game Tag a Tune **Verbosity** Squigl Matchin Fliplt PopVideo

Verbosity

it's common sense.

How to Play

- 1** You and a partner alternate between being **Describer** and **Guesser**. 
- 2** As **Describer**, you must help your partner guess the **secret word** by **giving clues**. 
- 3** As **Guesser** you must type the **secret word** that your partner is **describing**. 

Got it, Let's Play!

[View Top Scores](#)

One player gives clues about a word and the other player has to guess the word

clues

it is

it is a type of

it has

it looks like

about the same size as

it is related to

→ pass

Most Points Today

- 1 Catwoman **594 K**
- 2 Jeff **342 K**
- 3 PlasticBiddy **245 K**
- 4 jsm2530 **63 K**
- 5 You **47 K**
- 6 DaftyMcDaft **35 K**
- 7 Lottie **33 K**
- 8 guest228655 **11 K**
- 9 jMAC **9,250**
- 10 INTHEKY016 **8,300**

score
0



time
2:59

BONUS!
5,000 PTS ▶

the secret word is... shoe. 250 pts!

clues

- it is
- it is a type of
- it has
- it looks like
- about the same size as
- it is related to

→ pass

guesses



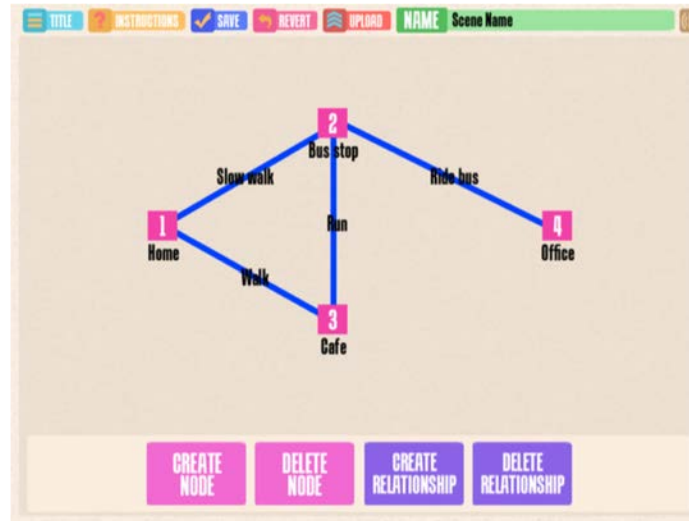
The clues written by players are learned as common sense statements

- Acceptance of statements:
 - Accept statements that are asserted more than once
 - Accept statements that appear first in the clue order
 - Exception: relation “it has” (frequently misused)
- Filters are applied:
 - Remove sound-alike clues (“*passion* looks like *fashion*”, “*wander* is a type of *wonder*”)
 - Remove clues with forbidden words (“letter”, “rhyme”, etc.)

The quality of the acquired knowledge is partially satisfactory

- Result of user evaluation:
 - Over 60% of collected sentences are generally true.
- Examples of wrong statements:
 - “Toy is a kind of *little*” (nonsense)
 - “Pearl is related to *shiny*” (bad grammar)

Gecka is a game engine for common sense knowledge acquisition



Designing video games helps to capture common sense knowledge

[Cambria et al., 2015] [Cambria et al., 2016]

Wikidata is a large scale database created by thousands of volunteers

- Goal of Wikidata:
 - Create a resource with easy access to data in a collaborative way similar to Wikipedia
- Characteristics of Wikidata:
 - Launched in October 2012
 - 22K active users (in 2019)
 - 72M items (in 2019)



Wikidata uses a frame-based representation for items

The image shows a Wikidata item page for 'Douglas Adams' (Q42) with various components annotated. The main label 'Douglas Adams' is linked to the item identifier '(Q42)'. Below the label is a description: 'English writer and humorist' and aliases: 'Douglas Noël Adams | Douglas Noel Adams'. The 'Statements' section is highlighted with a green box, labeled as a 'statement group'. It contains two main entries: 'educated at' and 'Brentwood School'. The 'educated at' entry is further detailed with a table of qualifiers: end time (1974), academic major (English literature), academic degree (Bachelor of Arts), and start time (1971). Below this is a section for '2 references', with one reference expanded to show details like 'stated in', 'reference URL', 'original language of work', 'retrieved', 'publisher', and 'title'. The 'Brentwood School' entry shows its own set of qualifiers (end time 1970, start time 1959) and a collapsed reference section.

label — Douglas Adams (Q42) — item identifier

description — English writer and humorist
Douglas Noël Adams | Douglas Noel Adams — aliases

property — educated at — value

rank —

statement group —

qualifiers —

opened references —

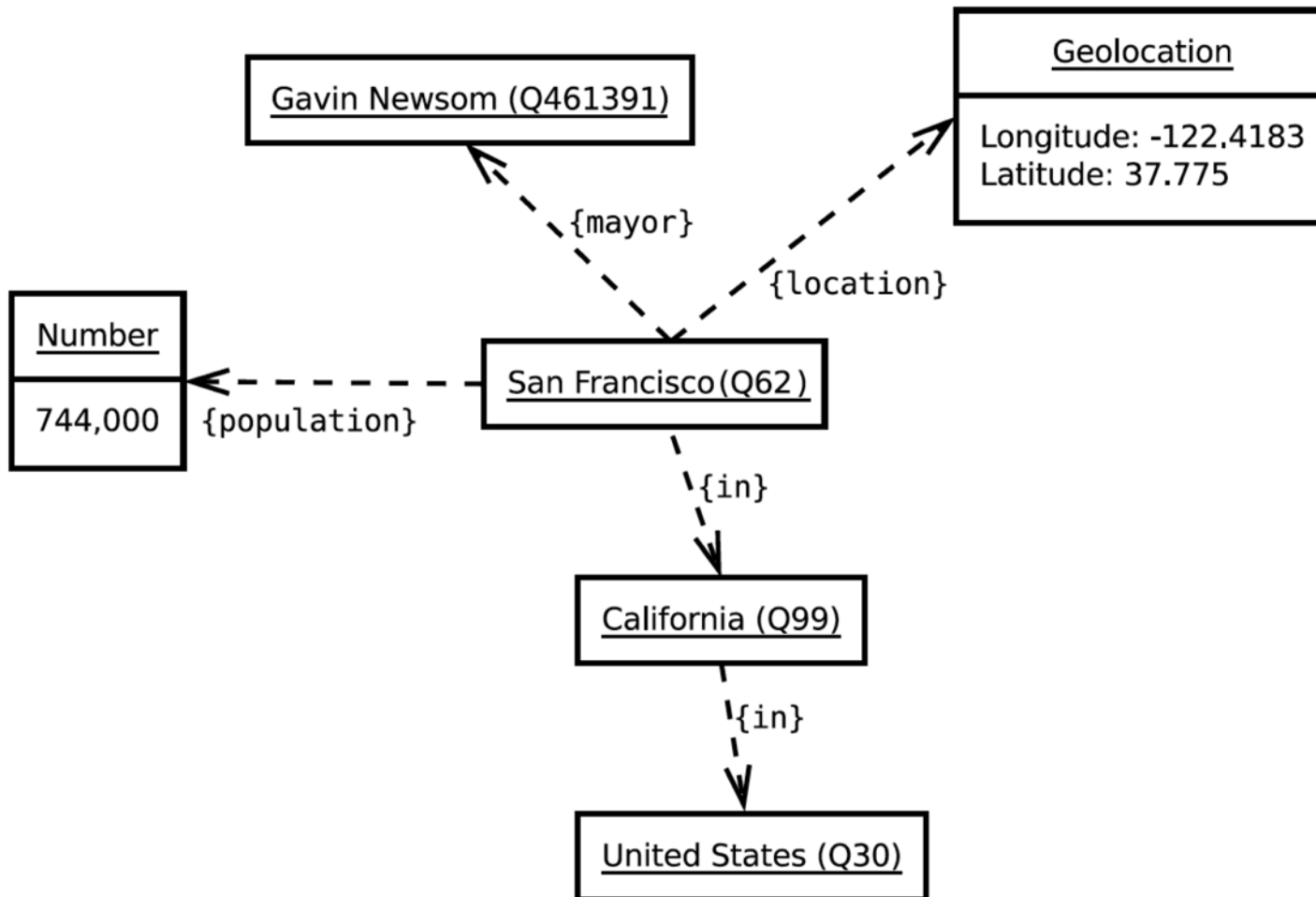
collapsed reference —

Property	Value
educated at	St John's College
end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971

Property	Value
stated in	Encyclopædia Britannica Online
reference URL	http://www.nndb.com/people/731/000023662/
original language of work	English
retrieved	7 December 2013
publisher	NNDB
title	Douglas Adams (English)

Property	Value
educated at	Brentwood School
end time	1970
start time	1959

Wikidata uses multiple relations between items



There are other initiatives for building collectively common sense knowledge bases

- Mindpixel
 - Web-based collaborative project to create a common sense knowledge base. The project was conceived by Chris McKinstry. Size: 1.4 million components (from 2000 to 2005).
- ISI's Learner
 - A web-based collaborative project to create a common sense knowledge bases (using analogy). Size: 600.000 facts (from 2003 to 2006). [Chklovski, 2003] <http://learner.isi.edu/>
- KBMC
 - A proposal using first-order probabilistic reasoning techniques to combine potentially inconsistent knowledge sources of varying quality [Richardson, Domingos, 2003]

What are the strengths and the weaknesses of collective knowledge acquisition?

- Strengths
 - The participation of thousands of volunteers reduces the cost and increases the speed of knowledge base development
 - Openmind emphasizes a content with universals (e.g., classes and relations between classes)
 - Wikidata emphasizes a content with facts (e.g., instances and relations between instances) with more than 70 million of items
- Weaknesses
 - Incomplete and unbalanced content
 - Knowledge can be biased by groups of interest
 - Less expressive representation than others (e.g., first order logic or probabilistic reasoning)

Course “Common sense reasoning”.
© 2019 Martin Molina

This work is licensed under Creative Commons license CC BY-NC-SA 4.0:
<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>



Work citation in APA style:

Molina, M. (2019). Common sense reasoning [Lecture slides]. OpenCourseWare, Universidad Politécnica de Madrid. Retrieved from <http://ocw.upm.es/course>