True Knowledge
Using AI to get Answers from the Internet

Simon Overell
ECIR Industry Day
Milton Keynes
1st April 2010
Using AI to get Answers from the Internet

• True Knowledge
  – Who are we
  – What we do

• How it works
  – System architecture
  – Data mining architecture

• Progress
  – Objects, facts, relations and NLP

• Get Involved
Who are we?

- Cambridge based start-up
- 24 people
- 3 1/2 years in business

Simon Overell

- Masters in AI and PhD in IR from Imperial College London
- Data Miner at True Knowledge
- Background in Search, NLP and Data Mining
What is True Knowledge?

An open-domain question answering platform

- **Semantic Technology**
  - Patented technology allowing True Knowledge to understand what a query means

- **Inference Engine**
  - Combined with strong semantics it is capable of answering trillions of queries including tail queries it has never seen before

- **Structured Knowledge**
  - *(continuously growing)* knowledge base of ~235 million structured facts

- **Integration with search**
  - Reliably answers when it knows, stays quiet when it doesn’t
How it works
Architecture

- Browser Interface
- Natural Language Translation
- Query/Answer System
- User Assessment
- System Assessment
- Knowledge Generator
- Knowledge Base
- External feeds
- Knowledge Addition
- Other Sources
Architecture
Architecture – Data Mining

Knowledge Base

Knowledge Addition

Static Mining
- UK Postcodes
- Biological Taxonomy

Live Mining
- Wikipedia
- Free Base
- Web Crawler
Static Mining Example

![True Knowledge](image)

*What would you like to know?*

What is the flight time from Barcelona to Luton?

*1 hour and 30 minutes*
Live Mining Example

What would you like to know?
List actors who died in 2010

James Mitchell
James Mitchell (born February 29, 1920 in Sacramento, California), the American actor and dancer, probably best known for his role as Palmer Cortlandt on the long-running soap opera All My Children
wikipedia

Ilan Carmichael
Ilan Carmichael OBE (born 16 June 1920), the English film, stage, television and radio actor
wikipedia

Howard Zinn
Howard Zinn (born 1922), the historian
wikipedia

Grethe Sønck
Grethe Sønck (16 July 1929 - February 12 2010), the Danish actress and singer
wikipedia
show all (56 more)
Ingredients

• Complete tree of (20k+) classes
  – Every entity is a member of multiple classes
  – Distinctness knowledge known

• Unique Recognition String (URS) for every entity

• Disambiguation techniques

• Simple knowledge representation

• Inference
How we represent knowledge (visual)

- Big Ben
  - was built in 1856
  - is located in London
  - Is an instance of City

- London
  - is located in The UK
  - Is a subclass of Geographical area

- The UK
  - can denote "united Kingdom"
  - Is an instance of Geographical area

- "united Kingdom"
  - Is an instance of Country

- Country
  - Is a subclass of Geographical area

- 1856
  - Is an instance of Time-point

ECIR 2010
Reality

• Basic KR is very simple
• Three* field database table:
  
  [london] [is an instance of] [city]

• Temporal knowledge represented by “facts about facts”
  
  [fact: [“123”]] [applies for timeperiod] [<1970 onwards>]

• Richness within entity representation
  – “parametered” objects
    • [integer: [“8128”]]
    • [group: [london]; [san francisco]]

* Actually 4 with negative relation
We use inference to extend knowledge

- Big Ben
  - was built in 1856
  - is located in London
  - is located in The UK

- London
  - is located in The UK

- The UK
Inference example

Who was the US president when Joe Biden was a teenager?

John F. Kennedy
Wikipedia page

Dwight D. Eisenhower
Wikipedia page

Your question
At the time when Joe Biden (born 1942), the American politician was a teenager (a human being between the ages of 13 and 19 inclusive) who was a US president, the president of the United States of America?

I used the following facts to provide this answer:
Joe Biden is a human being - agree / disagree
November 20th 1942 is the birthdate of Joe Biden - agree / disagree
John F. Kennedy has been a US president between January 20th 1961 and November 22nd 1963 - agree / disagree
Dwight D. Eisenhower has been a US president between January 20th 1953 and January 20th 1961 - agree / disagree

Show a detailed explanation
Progress
Successfully built Knowledge Base

Representation of the growth of both facts and objects in the True Knowledge Knowledge Base
Number of relations and classes

- Classes
- Relations
Ability to answer questions is improving
Traffic is continually growing

Monthly unique visitor growth
Jan. 2009 through Mar. 2010
Mining knowledge with NLP

- Scaling up populating our KB using uncurated facts extracted from unstructured text
- Standard NLP tools used (pos tagger, LR parsing etc.) plus the True Knowledge Platform
- TK provides disambiguation and common sense knowledge
  - Can throw away bad knowledge when contradicted by other things it knows

- **Achieving 98% accuracy** mining Wikipedia and News websites
- Facts also being mined from wider web
NLP mining of facts

Amur tigers, also known as Siberian tigers, live in the forests and mountains of Russia’s far east. This litter are the offspring of two adults transported from Edinburgh Zoo last October.

Iain MacDonald went to meet them.

[fact: ["auto4_3209433@trueknowledge.com"]]
[siberian tiger] [is a class of things doing the action location pair]
[group: [reside]; [area of forest]]

This fact asserts that the relationship ‘is a class with at least some members who regularly do the action/class of places pair’ exists between Amur Tiger and live and forest at some point in time. The additional fact [fact: ["auto4_3209434@trueknowledge.com"]] asserts that this fact applies for June 27th 2009 onwards.

Assessments
This fact has had the following assessments:
<table>
<thead>
<tr>
<th>Reporter</th>
<th>Source</th>
<th>Document</th>
<th>Time</th>
<th>Notes</th>
<th>Endorsement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[the [web_bot] unstructured text of bbc]</td>
<td>[url: [&quot;<a href="http://news.bbc.co.uk/2/hi/uk_news/8104260.stm">http://news.bbc.co.uk/2/hi/uk_news/8104260.stm</a>&quot;]]</td>
<td>2009-06-27 15:08:34</td>
<td>Fact extracted from BBC News using natural language processing of the following snippit “Amur tigers, also known as Siberian tigers, live in the forests and mountains of Russia’s far east.”</td>
<td>endorsed</td>
<td></td>
</tr>
</tbody>
</table>
Where do siberian tigers reside?

WikiAnswers - Why do tigers live in forest and jungle
Tigers how do they act in the forest? Where do tigers reside in the forest? Are there tigers in the decidous forest? Where is the tigers habitat in the jungle? Do siberian ...
wikipedia.com/Q/Why_do_tigers_live_in_forest_and_jungle  Cached page

WikiAnswers - How many types of tigers are there in the world and what ...  
Tigers question: How many types of tigers are there in the world and what are their names and where do they reside? ... SIBERIAN OR AMUR TIGER - the largest of the tiger subspecies ...
wiki.answers.com/Q/How_many_types_of_tigers_are_there_in_the_world_and_what_are_their...  Cached page

Tiger - Wikipedia, the free encyclopedia
The heaviest wild Siberian tiger on record weighed in ... Even apparently healthy white tigers generally do not live ... which allows tribal populations to reside inside designated tiger ...
en.wikipedia.org/wiki/Tigers  Enhanced view
NLP mining of facts

Ingmar Bergman dies at age 89

Monday, July 30, 2007

Swedish director Ingmar Bergman, a filmmaker widely regarded as one of the great masters of modern cinema, died today at his home on Fårö, Gotland, Sweden. He was 89. The exact cause of death is not known.

Biography

Born Ernst Ingmar Bergman in 1918 in Uppsala, Sweden to a Lutheran minister of Danish descent, Bergman grew up surrounded by religious imagery and discussion. He had a stint in an army at the outbreak of World War II, but became disillusioned with mandatory military service. He attended Stockholm High School and Stockholm University, without completing his degree in literature and art. He instead became interested in theatre and later in film.

...
NLP mining of facts

[事实："457057736@trueknowledge.com"]
[ingmar bergman] [is someone who studied at] [stockholm university]

这事实声明，在某关系'is someone who studied at'之间存在Ingmar Bergman和Stockholm University的关系，对所有时间。

undo  edit  agree  disagree


事实性质。

True（系统目前认为事实正确）
Uninfluenced（这个事实没有被其他事实所反驳，也没有被其他事实所取代）

这些性质在2009-09-01 10:31:39更新。

reassess this fact's properties

评估

这事实有以下评估：

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Source</th>
<th>Document</th>
<th>Time</th>
<th>Notes</th>
<th>Endorsement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[web_bot]</td>
<td>[unspecified member of [missing object class: fact pattern: [object unspecified]; [is responsible for content at: [domain name: [&quot;en.wikinews.org&quot;]]]]</td>
<td>[url: [<a href="http://en.wikinews.org/wiki/Ingmar_Bergman_dies_at_age_89">http://en.wikinews.org/wiki/Ingmar_Bergman_dies_at_age_89</a>]]</td>
<td>2009-09-01 10:31:36</td>
<td>Extracted by natural language processing from <a href="http://en.wikinews.org/wiki/Ingmar_Bergman_dies_at_age_89">http://en.wikinews.org/wiki/Ingmar_Bergman_dies_at_age_89</a> using snippet &quot;He attended Stockholm High School and Stockholm University, without completing his degree in literature and art&quot;</td>
<td>endorsed</td>
</tr>
</tbody>
</table>
Which university did Bergman go to?

Stockholm University
Stockholm University (Stockholms universitet), the state university in Stockholm, Sweden

Which are the universities (institutions of higher education) that Ingmar Bergman (born 1918), the Swedish director is a person who has studied at the educational establishment?

How do we know this?

I used the following facts to provide this answer:

- Ingmar Bergman is someone who studied at Stockholm University
  - agree
- Stockholm University is or has been a university
  - agree
The opportunity
Get Involved

YOUR COUNTRY NEEDS YOU.
Visit our main site

- [www.trueknowledge.com](http://www.trueknowledge.com)
- Ask Questions
- Add Knowledge
Find shops and products in your area

• Visit local.trueknowledge.com
Enhance your Google* Results

• Visit www.trueknowledge.com
• Install our IE or Firefox plugin

*And Bing, Yahoo and Ask
Build on our technology

• Register for an API account at http://www.trueknowledge.com/api/
• Visit our Forum and Blog for code examples
Questions?

simon.overell@trueknowledge.com

http://www.trueknowledge.com/