

Using Protégé for Automatic Ontology Instantiation

Harith Alani, Sanghee Kim, David Millard, Mark Weal,
Paul Lewis, Wendy Hall, Nigel Shadbolt



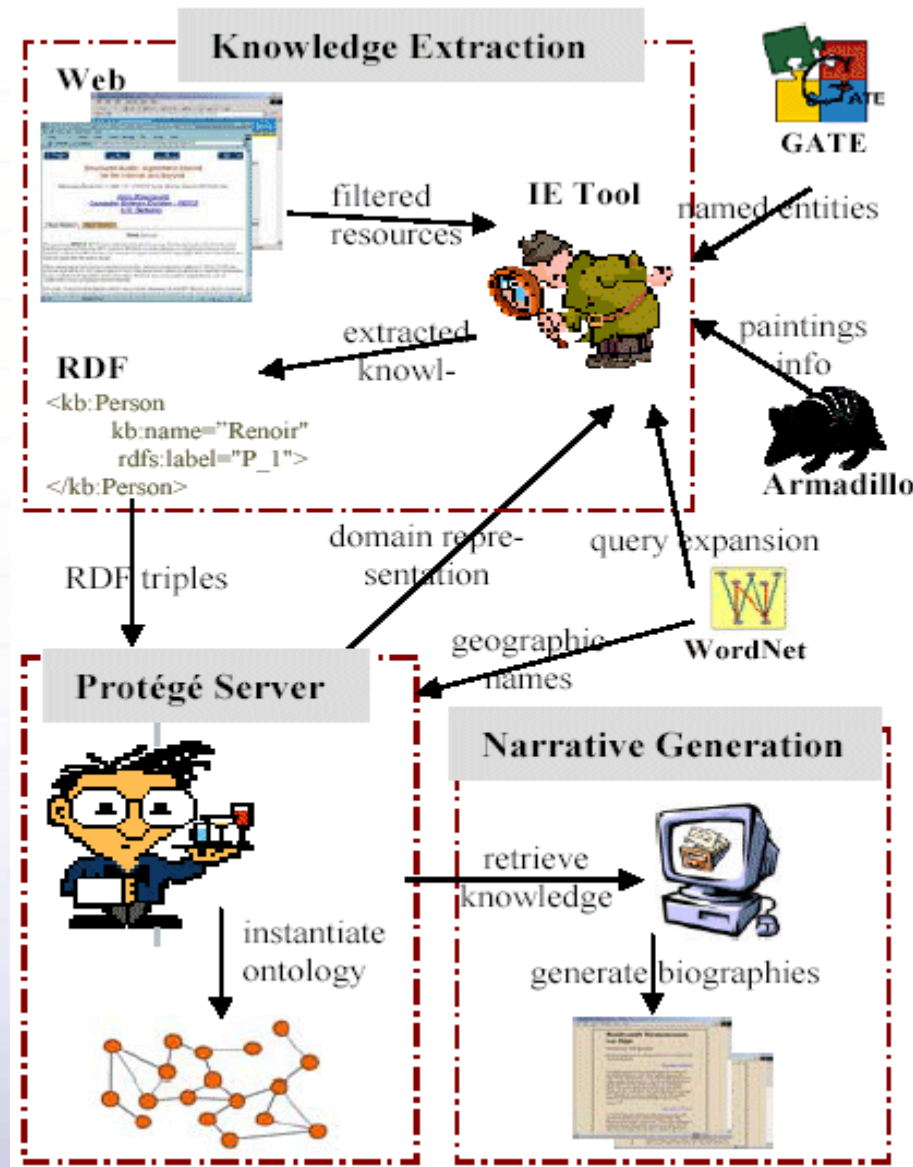
7th International Protégé Conference

- Aims:
 - Use NLT to automatically extract relevant information about the life and work of artists from online documents
 - Feed this information automatically to an ontology designed for this domain
 - Generate stories by extracting and structuring information from the knowledge base in the form of biographical narratives

Motivation

- The knowledge is out there!
 - Available on the web, buried in text documents, not understood by machines!
- Semantic annotation might help
 - Annotations are rare
 - In the near future, annotations will probably not be rich or detailed enough to support the capture of extended amounts of content
- Knowledge extraction
 - There will always be a need for tools that can locate and extract specific types of knowledge, and store it in a KB for further inference and use

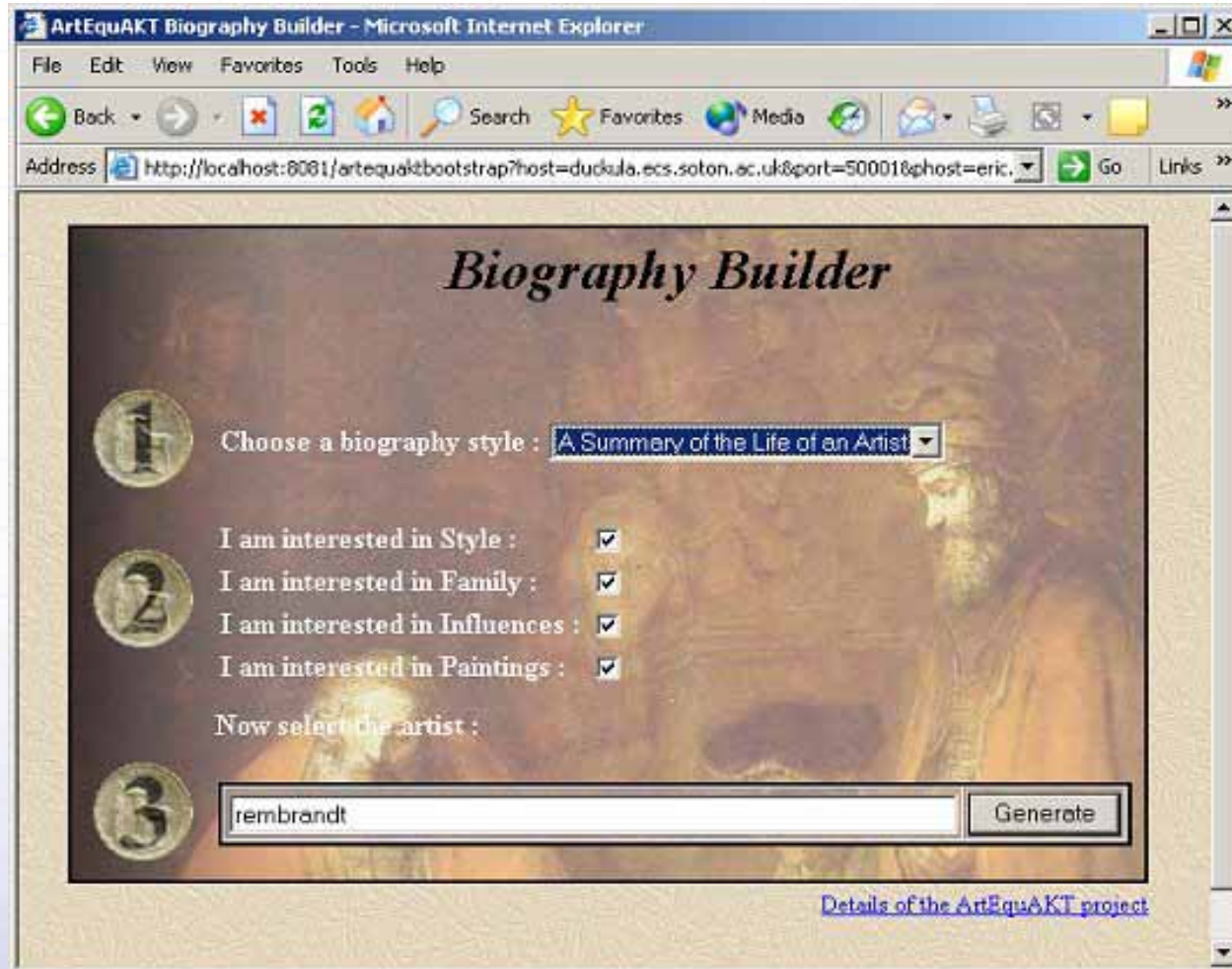
Architecture



ArtEquAKT Ontology

- Based on the Conceptual Reference Model (CRM) ontology
- Developed by CIDOC and promoted as an ISO standard
- CRM models the concepts and relationships used in cultural heritage documentation
- CRM is extended in ArtEquAKT to cover the life and work of artists

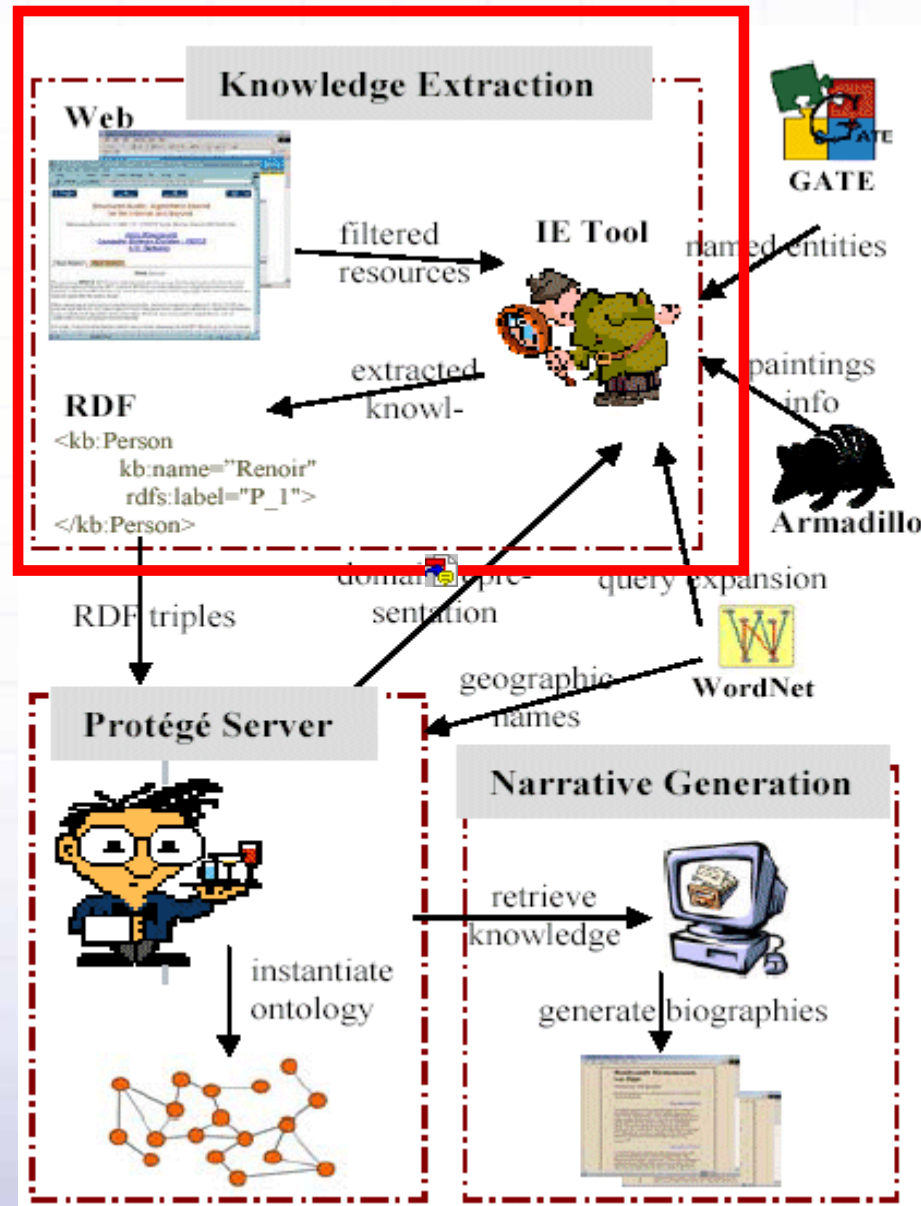
User Interface



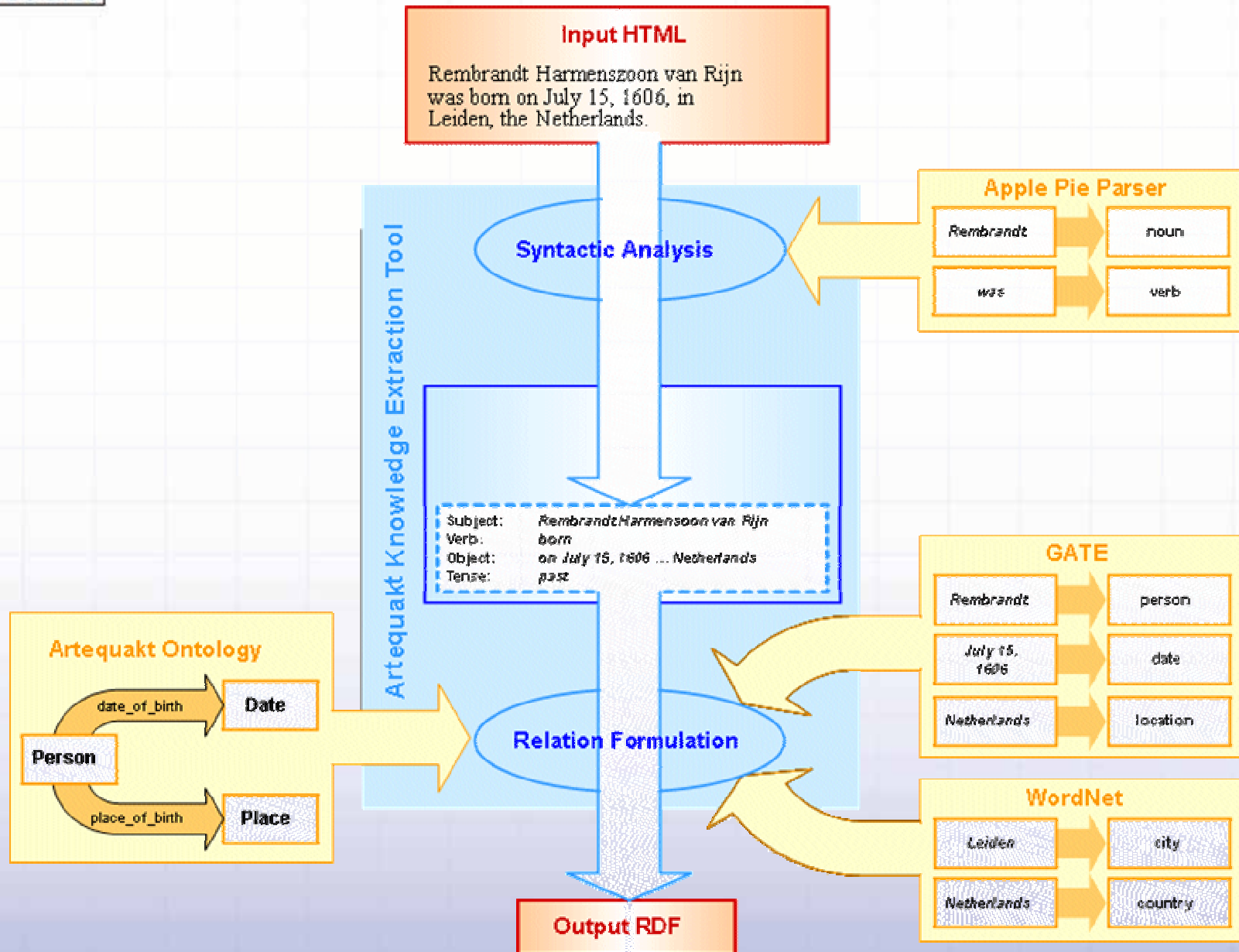
Search and Filter Documents

- Documents are selected following these steps:
 1. Query search engine (Google) with the given artist name
 2. Calculate the similarity of the returned documents to some example documents about artists
 3. Apply some heuristics (e.g. minimum paragraph length) to filter out documents containing mainly tables or hyperlinks
 4. Send the remaining documents to the information extraction process

Knowledge Extraction Component



Knowledge Extraction Process



Extraction Output

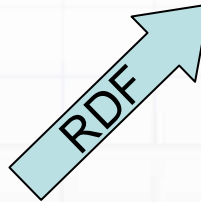
- Send the identified triples to the ontology server:

“Rembrandt Harmenszoon van Rijn was born on July 15, 1606, in Leiden, the Netherlands”

extracted triples

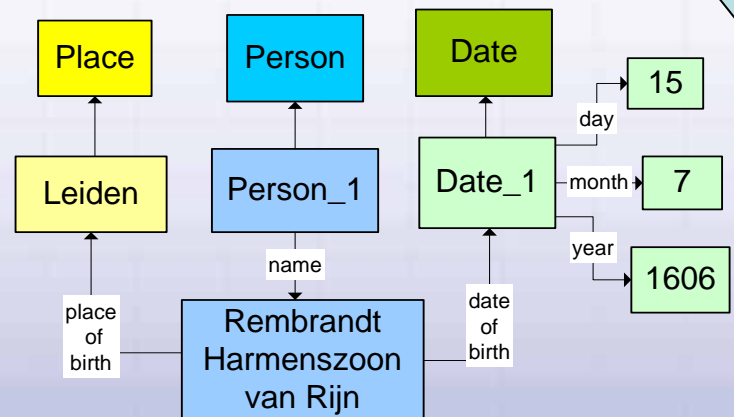
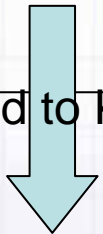


1. Person_1 *name* → Rembrandt ...
2. Person_1 *date_of_birth* → 15 July 1606
3. Person_1 *place_of_birth* → Leiden

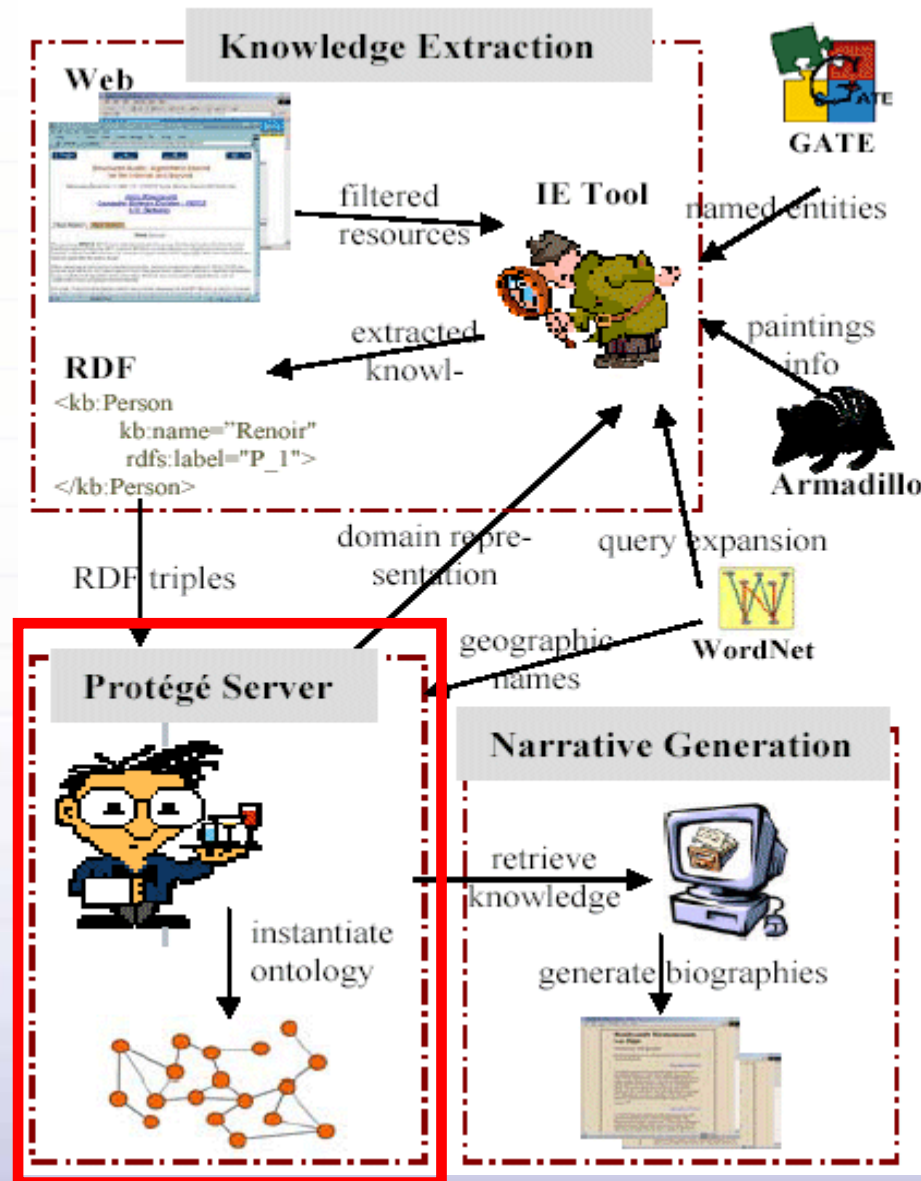


```
<kb:Person rdf:about="&kb;Person_1"
  kb:name="Rembrandt Harmenszoon van Rijn"
  rdfs:label="Person_1">
  <kb:date_of_birth rdf:resource="&kb;Date_1"/>
  <kb:place_of_birth rdf:resource="&kb;Place_1"/>
  <kb:has_information_text
    rdf:resource="&kb;Paragraph_1"/>
</kb:Person>
<kb:Date rdf:about="&kb;Date_1"
  kb:day="15"
  kb:month="7"
  kb:year="1606"
  rdfs:label="Date_1">
</kb:Date>
<kb:Place rdf:about="&kb;Place_1"
  kb:name="Leiden"
  rdfs:label="Place_1"/>
</kb:Place>
```

add to KB



Knowledge Management Component



Knowledge Management Process

- Provide guidance to the extraction process
- Receives extracted knowledge in RDF format
- Instantiate the ontology with the given knowledge triples (add to the KB)
- Consolidation the knowledge
- Verify inconsistencies
- Ontology server providing a set of inference queries

Knowledge Consolidation


Biography - Microsoft Internet Explorer

Address: <http://www.kfki.hu/~arthp/bio/r/rembrandt/biograph.html>

REMBRANDT Harmenszoon van Rijn
(b. 1606, Leiden, d. 1669, Amsterdam)

Pictures

Biography



Rembrandt was born in Leiden on July 15, 1606- his full name Rembrandt Harmenszoon van Rijn. He was the son of a miller. Despite the fact that he came from a family of relatively modest means, his parents took great care with his education. Rembrandt began his studies at the Latin School, and at the age of 14 he was enrolled at the University of Leiden. The program did not last long as he was expelled for his first painting, 'Swallowtail' (1624). He then moved to Amsterdam in 1631, where he came into contact with wealthy patrons who eagerly collected his work. An example from this period is the [Portrait of Nicolaes II de Vries](#).

WebMuseum: Rembrandt - Microsoft Internet Explorer

Address: <http://www.ibiblio.org/wm/paint/auth/rembrandt/>

WebMuseum, Paris


Rembrandt

Rembrandt HARMENSZOOON VAN RIJN (b. July 15, 1606, Leiden, Neth.--d. Oct. 4, 1669, Amsterdam), Dutch painter, draftsman, and etcher of the 17th century, a giant in the history of art. His paintings are characterized by luxuriant brushwork, rich colour, and a mastery of chiaroscuro. Numerous portraits and self-portraits exhibit a profound penetration of character. His drawings constitute a vivid record of contemporary Amsterdam life. The greatest artist of the Dutch school, he was a master of light and shadow whose paintings, drawings, and etchings made him a giant in the history of art.

Rembrandt Harmenszoon van Rijn was born on July 15, 1606, in Leiden, the Netherlands. His father was a miller who wanted the boy to follow a learned profession, but

REMBRANDT - Microsoft Internet Explorer

Address: <http://home.swipnet.se/~w-17719/Index.htm>

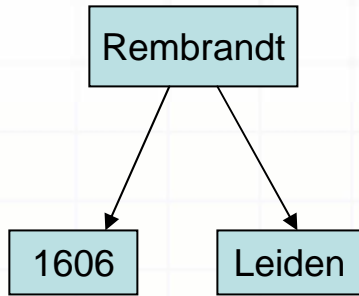


- INFO
- PAINTINGS
- LINKS
- E-MAIL

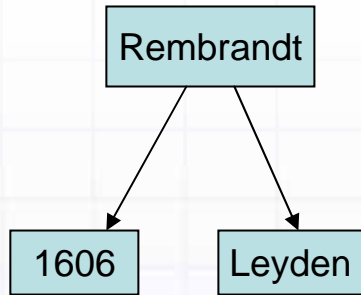
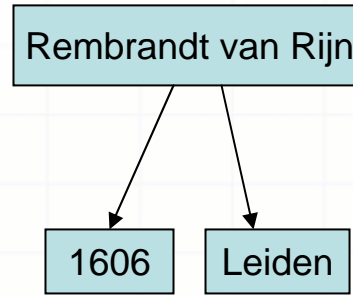
Welcome to my site about Rembrandt Harmenszoon van Rijn.

Rembrandt was born in Leyden in 1606, the son of a miller. He died sixty-three years later in Amsterdam. He was an artist of prodigious output, there are in the neighborhood of 650 paintings by him, about 300 etchings and between 1,500 and 2,000 drawings. Here you'll find some information about Rembrandt, and a few of his amazing paintings and a few links to other sites about him as well. I hope you will find this site interesting, and you are welcome to mail me and praise/flare me about it...

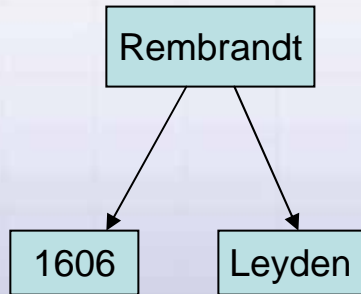
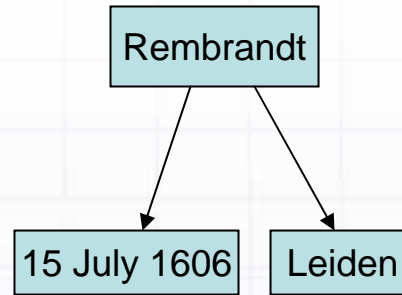
Types of Duplication



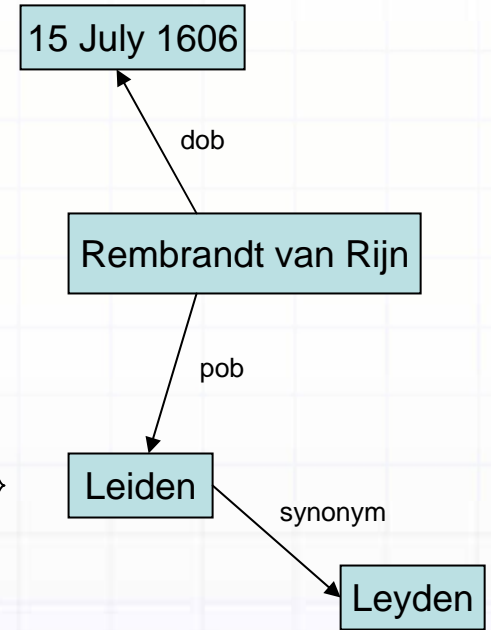
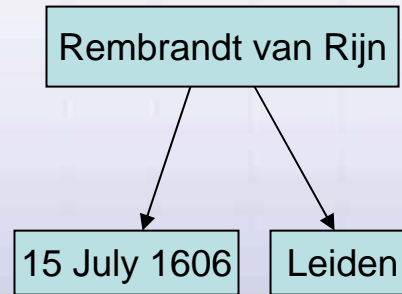
duplicate instances of the same artist



duplicate attribute values



duplicate instances *and* attribute values



Consolidation Procedure

- Unique Name Assumption
 - e.g all “Rembrandts” are merged
 - Not fool-proof, but works well in this limited domain
- Information Overlap
 - Merge similarly named artists if they share specific attribute values
 - e.g. Rembrandt, and Rembrandt Harmenszoon share a date of birth and a place of birth
- Merge less specific information into more detailed ones
 - This is mainly performed for dates and places
 - e.g 1606 is merged into 15/7/1606; Netherlands is merged into Leiden
 - Place names are expanded with WordNet
 - Synonyms: Leiden = Leyden
 - Holonyms (part of): Leiden is part of The Netherlands
 - What if there is more than one Leiden? How do we know which to select?
 - Use the specificity variation of the given place for disambiguation
 - e.g. we are here looking for a Leiden that is related to the Netherlands

Verifying Inconsistencies

WebMuseum: Holbein, Hans the Younger - Microsoft In...

Address: <http://www.ibiblio.org/wm/paint/auth/holbein/>

Google: webmuseum hans holbein

Hans Holbein the Elder (1465?-1524). Born in Augsburg, Bavaria, the elder Holbein, like his brother Sigmund, painted richly colored religious works in the late Gothic style. In addition to the [altar paintings](#) that are his principal works, he designed church windows and also made a number of portrait drawings that foreshadow the work of his famous son. His later paintings show the transition from the late [Gothic](#) to the [Renaissance](#) style. He died in Isenheim, Alsace.

Hans Holbein, the Elder -- Encyclopædia Britannica - Microsoft Internet Explorer

Address: <http://www.britannica.com/eb/article?eu=41661>

Google: hans elder

Holbein, Hans, The Elder
Encyclopædia Britannica Article

Article

[E-mail this article](#) [Print this article](#) [Cite this article](#)

born c. 1465, Imperial Free City of [Augsburg](#) [Germany] died 1524, Isenheim, Alsace [now in France]

German painter associated with the Augsburg school. He was the senior member of a family of painters that included his brother

Wikipedia - Microsoft Internet Explorer

Address: http://en.wikipedia.org/wiki/Hans_Holbein_the_Elder

Hans Holbein the Elder
From Wikipedia, the free encyclopedia.

Hans Holbein (1460 - 1524) was a [German painter](#).

Hans Holbein the Elder (around 1460/5-1524) was born in [Augsburg, Bavaria](#) and died in Isenheim, Alsace. He and his brother Sigmund Holbein painted religious works in the late Gothic style. Hans the Elder was a pioneer and leader in the transformation of German art from the Gothic to the Renaissance style.



[Discuss this page](#)
[Page history](#)
[What links here](#)
[Related changes](#)

[Special pages](#)
[Contact us](#)
[Donations](#)

Address: http://en.wikipedia.org/wiki/Image:Hans_Holbein.jpg


Olga's Gallery - ...

Address: cgallery.com/H/holbein/

Gallery

Hans Holbein the Elder
(1460-1524)

[Biography](#)



Death of the Virgin. c. 1501.
Tempera on wood.
Kunstmuseum Basel, Basel, Switzerland.
[More.](#)

Verifying Inconsistencies

- We don't aim for "the right answer", but for some sort of a confidence value
- But which answer is *more likely* to be the correct one?
 - *Trust*: certain sources can be more trusted than others, but how do we judge that?
 - *Frequency*: certain facts might be extracted more often than others
 - *Extraction*: some extraction rules are more reliable than others!

Instantiated Ontology

Artequakt5-WS04 Protégé 2.0.1 (C:\Documents and Settings\ha\Desktop\Artequakt5-WS04.ppr, Standard Text Files)

Project Edit Window Help

Classes Slots Forms Instances Queries

Classes

- THING
- SYSTEM-CLASS
- E1.CIDOC_Entity
 - E18.Physical_Entity
 - E19.Physical_Object
 - E22.Man-Made_Object
 - E23.Iconographic_Thing
 - Painting
 - E26.Physical_Feature
 - E2.Temporal_Entity
 - E28.Conceptual_Object
 - E36.Visual_Item
 - Information_Text
 - Document (46)
 - Paragraph (374)
 - Sentence (2519)
- E39.Actor
 - E40.Legal_Body
 - Person (19)
 - E52.Time-Span
 - E53.Place (535)
 - E54.Dimension
 - E55.Type
 - E59.Primitive_Value
 - E61.Time_Primitive
 - Extracted_Triple (9883)
 - Study_information (14)
 - Work_information (139)

Display Slot

name

Direct Inst:

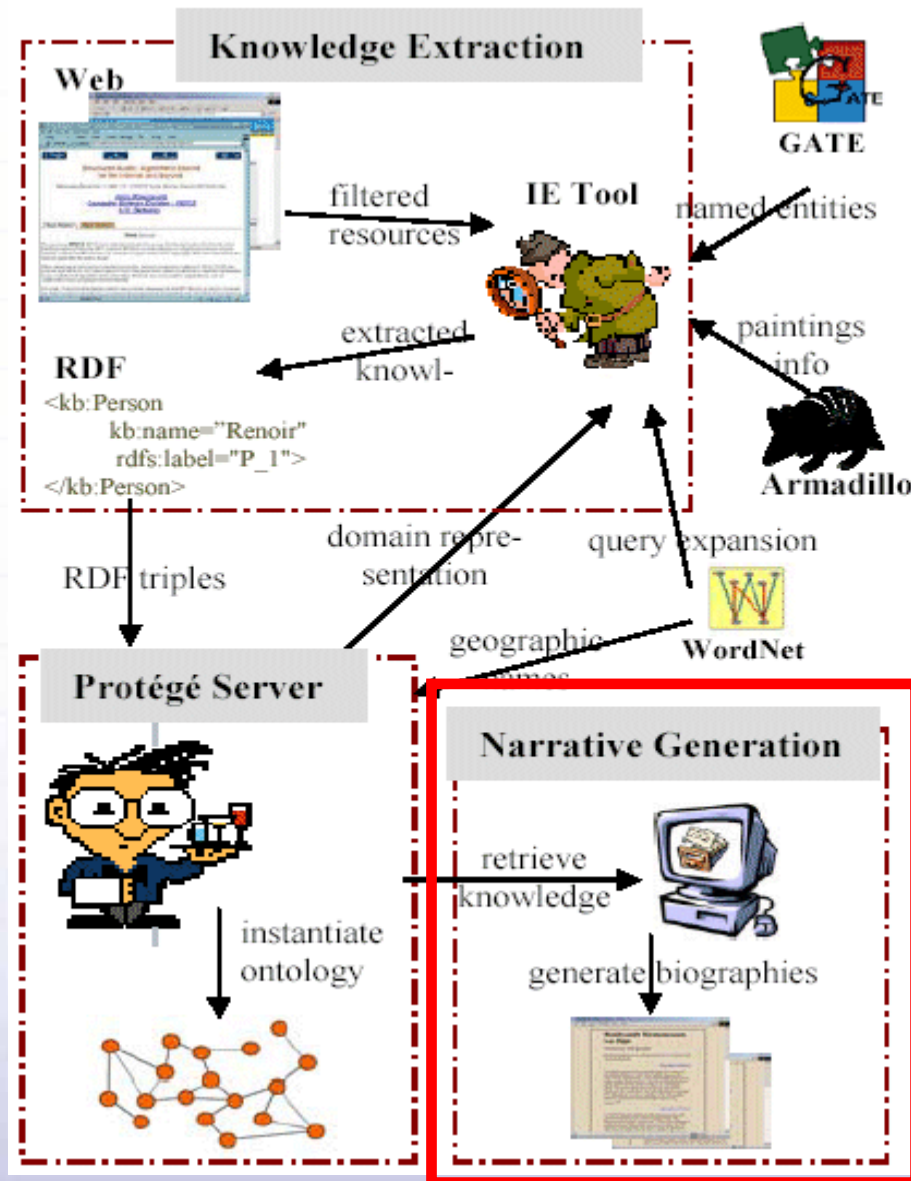
- <Person_105 name not set>
- <Person_108 name not set>
- <Person_13 name not set>
- <Person_132 name not set>
- <Person_133 name not set>
- <Person_26 name not set>
- <Person_27 name not set>
- <Person_43 name not set>
- Alfred Sisley
- Francisco Jose de Goya
- Gustave Courbet
- Mary Cassatt
- Mary Stevenson
- Pierre Auguste Renoir**
- Rembrandt Harmenszoon van Rijn
- Rembrandt Peale
- Saskia van Uylenburgh
- Saskia van Uylenburgh
- Sisley

Pierre Auguste Renoir (type=Person, name=Person_63)

| | |
|-----------------------|--|
| Name | Painted |
| Pierre Auguste Renoir | <ul style="list-style-type: none"> a Tuke Henry a Tissot James a Sargent Singer a Delacroix Eugene |
| Date Of Birth | Place Of Birth |
| 25/2/1841 | France |
| Date Of Death | Place Of Death |
| 3/12/1919 | France |
| Date Of Marriage | Married To |
| //1890 | |
| Painting Style | P21.1F.is Member Of |
| | |



Narrative Generation Component

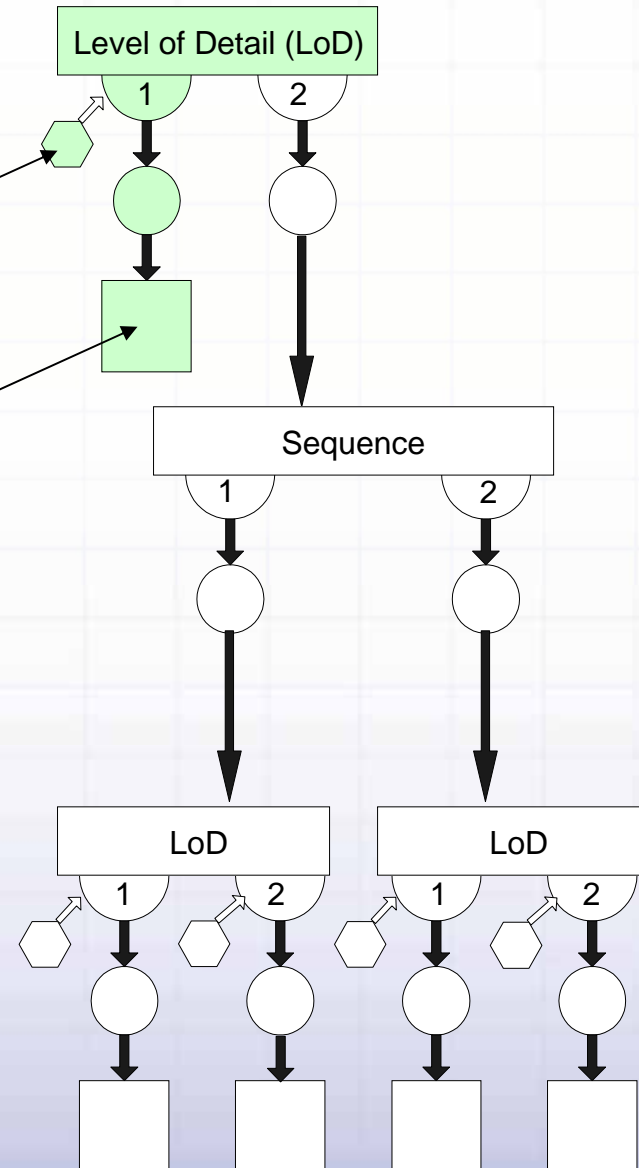


Narrative Generation

Intro paragraph :
 DOB + place

Paragraph with DOB and Place

Rembrandt Harmenszoon van Rijn was born on **July 15, 1606**, in **Leiden, the Netherlands**. His father was a miller who wanted the boy to follow a learned profession, but Rembrandt left the University of Leiden to study painting.



Best option is to have one paragraph that contains both pieces of information

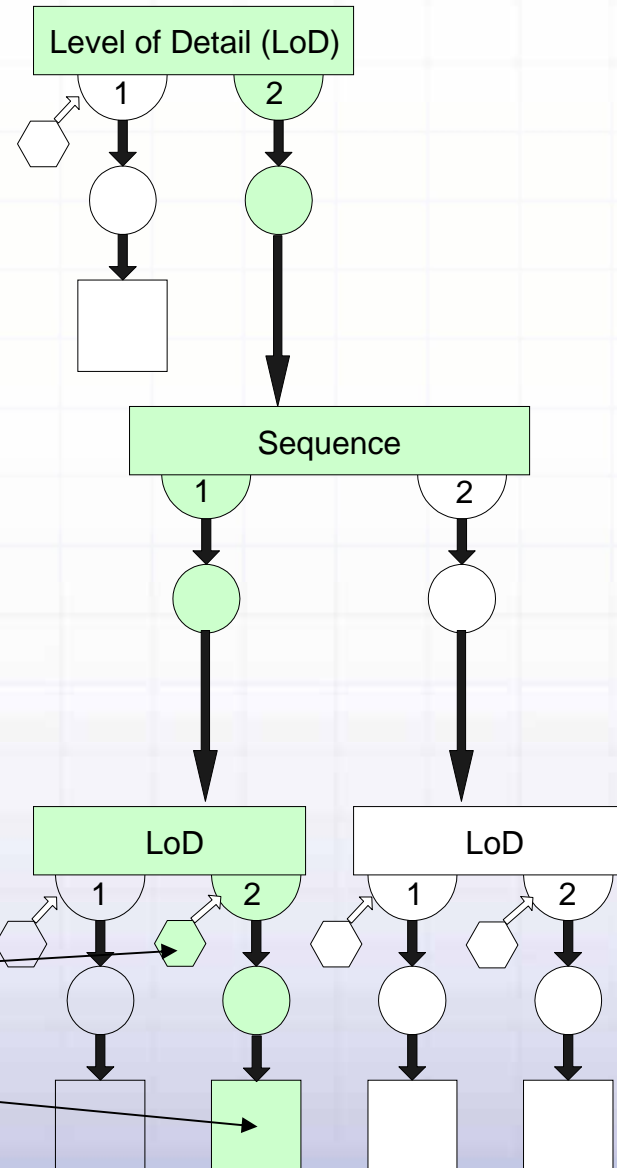
FOHM Template

Intro paragraph :

DOB + place

Otherwise need a sequence of two fragments (DOB and place).

Either use a paragraph for each fragment, or construct out of raw facts



Constructed sentence:
Rembrandt was born on **July 15, 1606**.

Rembrandt Harmenszoon van Rijn

Summary Biography

Rembrandt Harmenszoon van Rijn was born on 15 July 1606 in Leiden, Netherlands.

[More detail available \(2\).](#)

Rembrandt married Saskia van Uylenburgh, the cousin of a successful art dealer, in 1634. The marriage enhanced his career and brought him into contact with wealthy patrons who eagerly sought out his services. She had a substantial fortune, and this allowed Rembrandt to indulge in an extravagant lifestyle. He spent a great deal of money on an extensive collection of art. At auctions, he would often place an opening bid so high that no one else could place a bid. He bought a large house that later would place further strain on his resources.^[1]

[More detail available \(2\).](#)

In 1625 Rembrandt returned to Leiden and set up his own independent studio and in 1629 he began the first of his numerous self-portraits (Self Portrait , 1929). The paintings he completed in his Leiden studio show a more mature use of chiaroscuro technique than his master, Lastman. Instead of using the lighting effect to create drama, Rembrandt used it to compose the entire painting. His use of light and dark gave his subjects a physical presence which for the first time involved the viewer in the painting. In his Leiden studio Rembrandt worked alongside Jan Lievens (1607 - 74) also a former student of Lastman. The two often worked on identical themes. In 1929 when Rembrandt produced paintings focusing on the apostle Paul (Peter and Paul in Conversation), Lievens painted the same apostle.^[4]

Rembrandt Harmenszoon van Rijn, Dutch baroque artist, painted The Return of the Prodigal Son shortly before his death in 1669. It now hangs in the Hermitage in St. Petersburg, Russia. The subject comes from the Gospel According to Luke , Chapter 15. Gaze upon this painting and you will appreciate a sense of great tragedy. The contrast between the light and the dark, the two mysterious figures that appear dimly in the background, the woman behind the father, the stoic appearance of the old man who stands on the side, and the dishevelment of the younger brother, all contribute to this sense of tragedy. One author has said that Rembrandt has elevated tragedy to a symbol of universal significance in this painting. I couldn't agree more. This image represents the epitome of Rembrandt's psychological mastery.^[1]

Rembrandt was the first artist to practice self-portraiture as a speciality. In doing so, he created a medium for self-fashioning that has since inspired many artists. In 1968 a committee of Dutch art historians known as the Rembrandt Research Project took upon itself to distinguish between real Rembrandt paintings and all others. The project has met with less acceptance than anticipated.^[6]

Rembrandt Harmenszoon van Rijn died 4 October 1669 in Amsterdam.

[More detail available \(2\).](#)

| Style : | <input checked="" type="checkbox"/> | Ref | Occ | Url |
|---|-------------------------------------|-----|-----|---|
| Family : | <input checked="" type="checkbox"/> | [1] | 3 | http://www.tiu.edu/psychology/rembrandt.htm |
| Influences : | <input checked="" type="checkbox"/> | [2] | 1 | http://www.ibiblio.org/wm/paint/auth/rembrandt/ |
| Paintings : | <input checked="" type="checkbox"/> | [3] | 1 | http://www.kfki.hu/~arthp/bio/r/rembran/biograph.html |
| References : | <input checked="" type="checkbox"/> | [4] | 3 | http://www.theartgallery.com.au/arteducation/greatartists/rembrandt/about/ |
| Paragraphs : | <input checked="" type="checkbox"/> | [5] | 1 | http://www.londonfoodfilmfiesta.co.uk/artmai~1/rembra~1.htm |
| <input type="button" value="Regenerate"/> | <input checked="" type="checkbox"/> | [6] | 1 | http://www.mystudios.com/art/bar/rembrandt/rembrandt.html |

Example Biography

ArtEquAKT Challenges

- **Extraction**
 - Some fact are too complex to extract
 - Rule based IE is not always sufficient
 - Mapping of ontology terms to those in the text is unreliable (better for the ontology editor to include synonymous terms)
- **Generation**
 - A much wider range of facts should be extracted to be able to generate the biographies from scratch
 - Narrative construction may require richer semantic support (e.g. ontology of narrative)
 - Generation is not error free. We rely on people's ability to parse and understand text
 - Difficult to track what facts has been included in the biography if these facts have not bee identified
- **Consolidation**
 - Unreliable if the facts are extracted incorrectly
 - Could be inaccurate with spars information
 - Geographical expansion can be wrong for places with same names
- **Planning a bid for a second generation of ArtEquAKT**
 - Entirely ontology driven
 - Domain independent
 - Much better text generation

Questions you may want to ask!

1. So does this system work with other domains?
2. Why bother with biographies anyway! There are many out there already!
3. Why extract knowledge, then use whole paragraphs in your biographies?!
4. Did you evaluate any of this?
5. What kind of knowledge did you manage to extract?
6. What did you say that Armadillo thing does?
7. How can we get GATE to recognise different entities?
8. How much rubbish does your system extract?
9. Can we use this system?! please?

10. How would you like me to fund you? cash or check?