

# Automatic Process Model Discovery from Textual Methodologies

**Elena V. Epure**

Patricia Martín-Rodilla

Charlotte Hug

Rebecca Deneckère

Camille Salinesi

# Who are we?



Patricia



Charlotte



Rebecca



Camille

# Introducing Dan



<http://www.artofmanliness.com/2009/11/11/so-you-want-my-job-archaeologist/>

« Archaeology is a **process**, and a very **methodical** one at that. »



« Archaeology is a destructive science »



« Managing an active dig site is like trying to keep **multiple plates spinning at once.** »



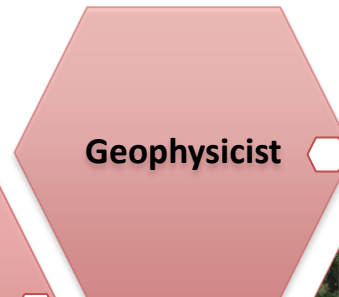




**Photographer**



**Technicians**



**Geophysicist**



**Drawers**



**Excavation  
team**

# Introducing Villa Magna



Home

## Welcome to Villa Magna

The Villa Magna Project aims at the investigation by excavation and survey of a large imperial Roman villa known from letters of Marcus Aurelius and its estate, and the subsequent life of the site, its fortification in late antiquity and the creation of a monastery among the ruins in the 10th century. The project is sponsored by the University of Pennsylvania's Museum of Archaeology and

## LATEST UPDATES

The most recently updated articles for the Villa Magna Project.

---

National Geographic special features Villa Magna

---



# Experiences are all around us

amazon.com

Annual  
Reports

To our shareowners:

As regular readers of this letter will know, our energy at Amazon comes from the desire to impress customers rather than the zeal to best competitors. We don't take a view on which of these approaches is more likely to maximize business success. There are pros and cons to both and many examples of highly successful competitor-focused companies. We do work to pay attention to competitors and be inspired by them, but it is a fact that the customer-centric way is at this point a defining element of our culture.

One advantage – perhaps a somewhat subtle one – of a customer-driven focus is that it aids a certain type of proactivity. When we're at our best, we don't wait for external pressures. We are *internally* driven to improve our services, adding benefits and features, before we have to. We lower prices and increase value for customers before we have to. We invent before we have to. These investments are motivated by customer focus rather than by reaction to competition. We think this approach earns more trust with customers and drives rapid improvements in customer experience – importantly – even in those areas where we are already the leader.

# Experiences are all around us

**Elena V. Epure,**

The latest information from [Introduction to Data Science](#) by [University of Washington](#)



Emails

Quick update on recent content:

Last week, we released new lectures on NoSQL with a goal of giving an overview of what's going on in the space and how the different buzzwords and systems fit together.

We also released an optional AWS assignment that gives you the opportunity to process a large graph dataset consisting primarily of social network information.

Yesterday, we released some lectures on selected topics in statistical analysis. Our goal here was not to try and give a compressed treatment of a stats 101 course -- besides not being feasible in the time allotted, we thought this approach would be a bit dry. Instead, we tried to identify topics that are not typically covered in early statistics courses, but should be.

The assignment we will release today or tomorrow will involve participating in a Kaggle competition; your peers will review an English description of your solution for

# Experiences are all around us

CAN SUPPLIERS CLAIM  
THAT THEIR  
GOODS OR SERVICES  
MEET ISO 9001?

No. The reference to ISO 9001 indicates that the supplier has a **quality management system** that meets the requirements of ISO 9001:2008. As mentioned earlier, this should provide you with confidence in your supplier's **ability** to provide consistent, conforming goods or services. ISO 9001 requires your supplier to monitor the levels of satisfaction of its customers (this includes you!), and to feed back this information in order to improve the effectiveness of its QMS.

WH  
IF THING

ISO  
Standards

In the event you are not happy with specific goods or services you receive, you should first of all bring this to your supplier's attention. You will typically do this via the normal technical and/or commercial communication channels that have been established. Your supplier is obliged to investigate your complaint and should take appropriate actions to avoid or reduce the chances of it happening again.

If, however, you are dissatisfied with the **overall** performance of your supplier (for example if they continue to provide nonconforming products, do not address your complaints, or are not taking appropriate corrective action), then this is an indication of problems in their **quality management system**. Depending on the responses you receive, you should be aware

# Experiences are all around us

## Quick Start

[Wiki HOWTOs](#)[Stable HOWTOs](#)[Mirrors](#)[Mailing lists](#)

Documentation

**warning** The LDP suffers of many spam attacks, by creation of dummy wiki accounts only to post spam. If you want to create a new account, please subscribe the discuss list and announce your intention there, if not you are at risk of being blocked. So!

## The Linux Documentation Project Wiki

The Linux Documentation Project - LDP for short, or tldp (taken from the domain name) - is a very long-standing project whose aim is to provide documentation for Linux software. The Project is governed by the [LDP Manifesto](#) and has a long [history](#).

The LDP website consists of two parts: the first part is a collection of documents, mostly the famous "HOWTO"s, but also some Gu source for rendering in [many other document formats](#). This part of the site is [mirrored all over the world](#).

The second part is the wiki you are currently reading. This part aims to be the LDP workshop, where new documents are written, and all are allowed to do so; but others can use the power and friendliness of this wiki to produce their texts. Be warned, however, that docs on

## LDP Users (Visitors, Authors, Members)

If you are reading this, you are already an LDP user 😊. Anyone can be a visitor if you only want to peruse our docs; but you can send mail to warn when something is wrong on the LDP Web site. You can be an author if you want to be responsible for a page, see the [author page](#). You can become a member - see the relevant page on the blue inset on the top right of any page. Thanks!



# Experiences are all around us




Medical  
Narratives

## NARRATIVE MEDICINE *HONORING THE STORIES OF ILLNESS*

FRIDAY, FEBRUARY 15, 2013  
10:00 AM TO 1:00 PM ATRIUM AMPHITHEATER

KEYNOTE SPEAKER  
*RITA CHARON, MD, PhD*

*SPECIAL GUESTS*  
*JAMES M. STUBENRAUCH, MFA*  
*JOY JACOBSON, MFA*



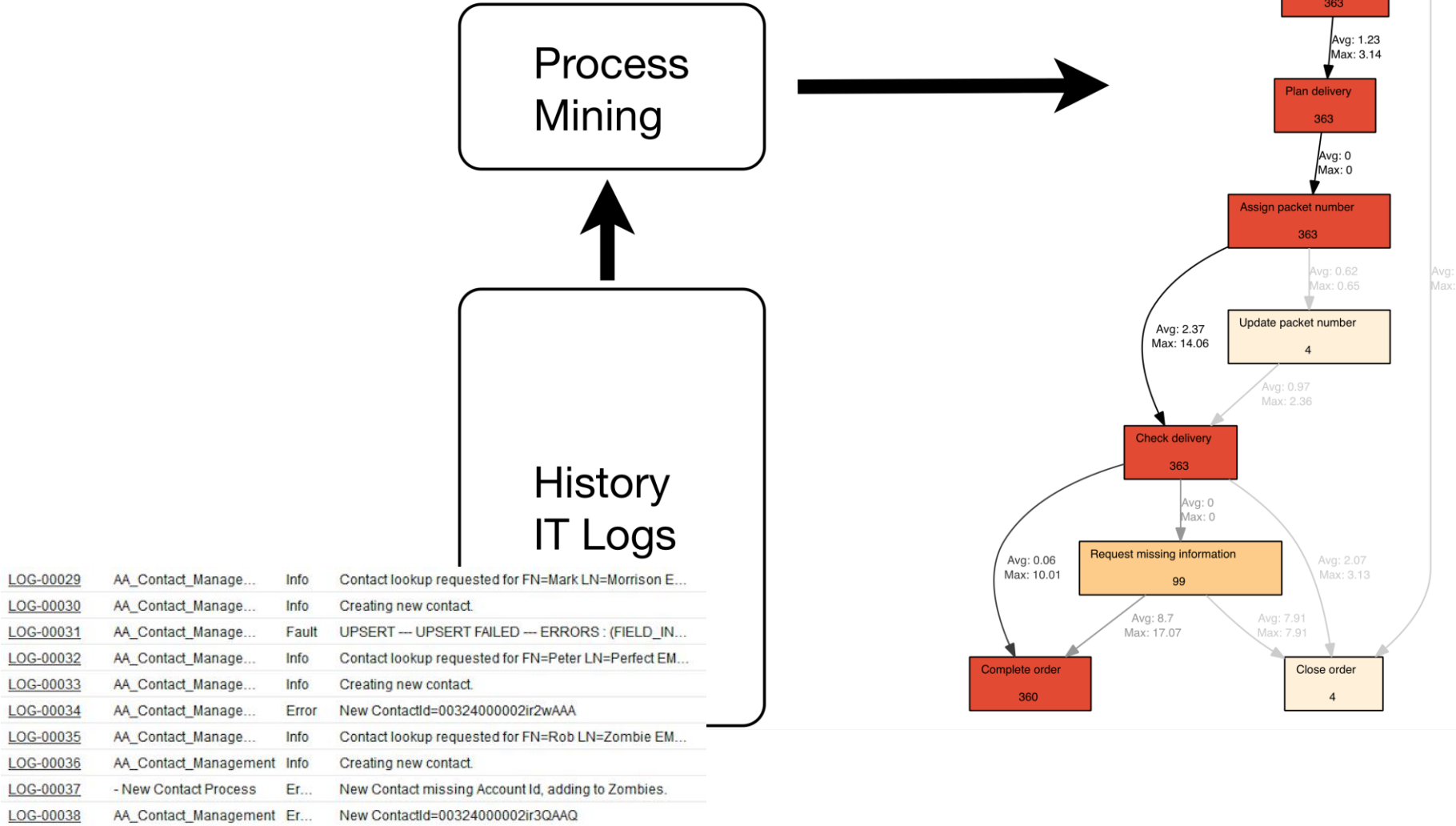
*COMPARATIVE  
PERSPECTIVES  
ON HEALTH, ILLNESS  
AND HEALING SERIES*



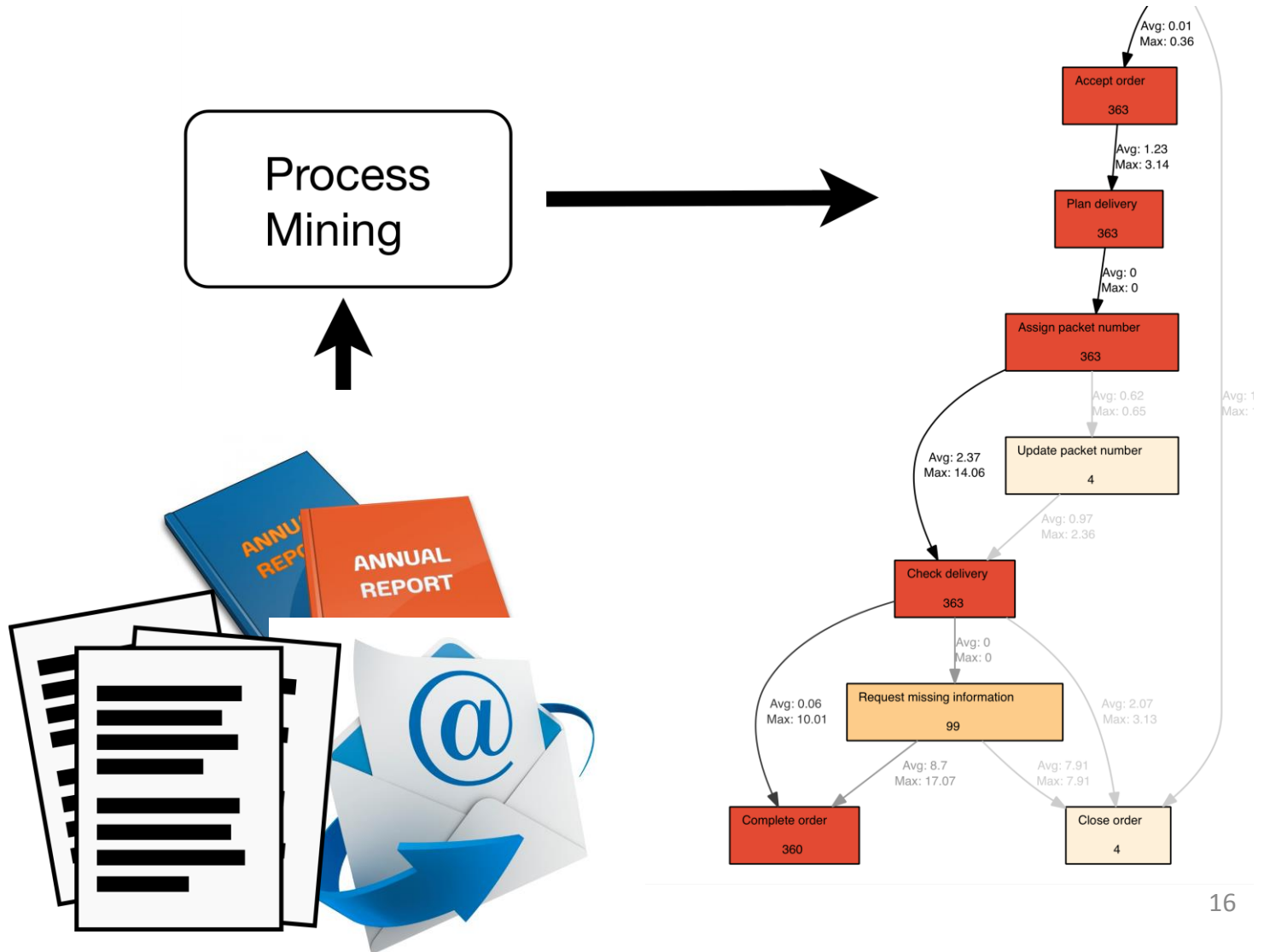
# Outline

- Introduction
- Related works
- Solution
- Validation in Archaeology
- Conclusion

# Context



# How to use natural language sources for producing structured knowledge as process instance models?



# Definitions



# Challenges

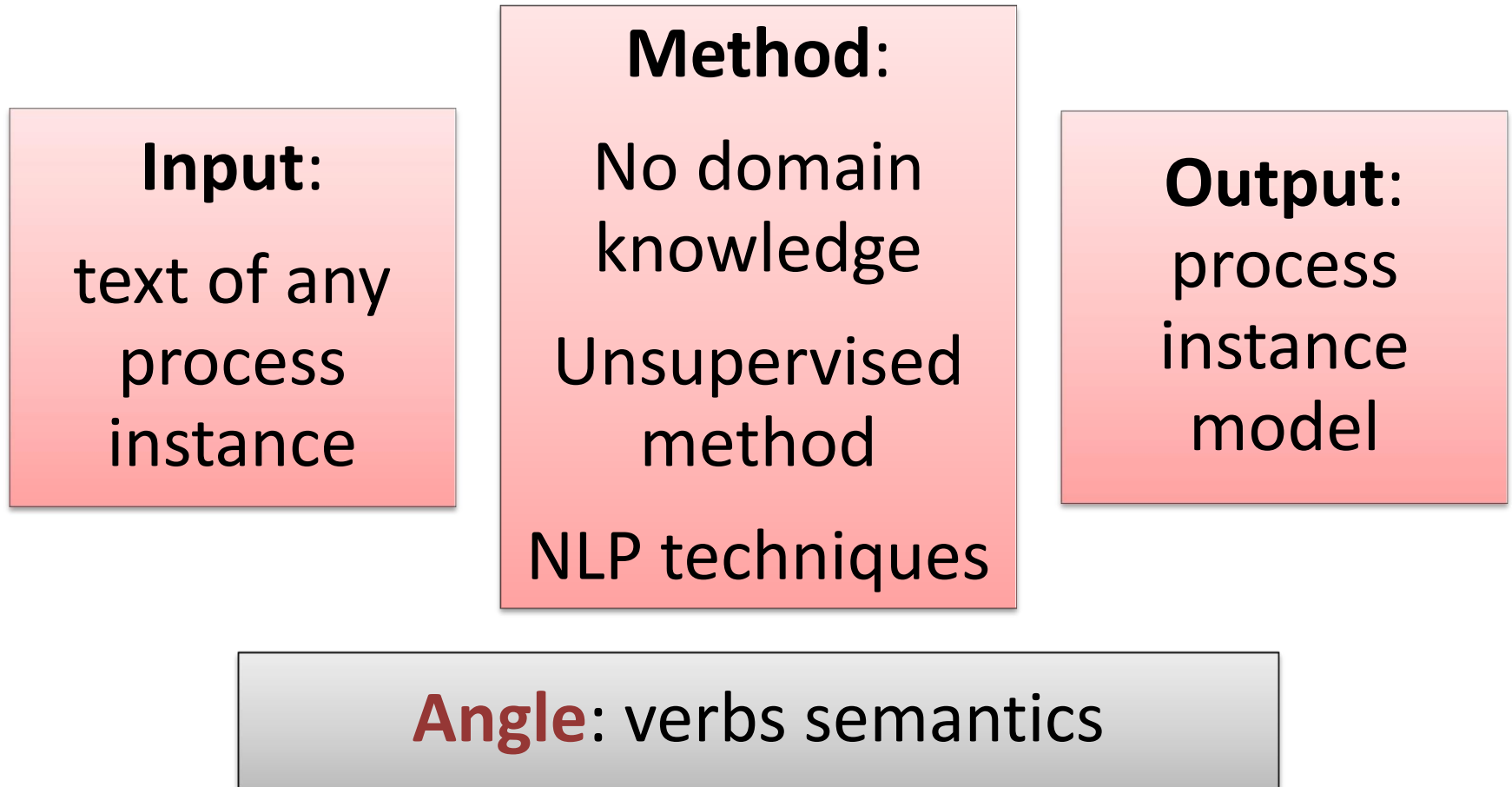
- Lack of structure
- Ambiguity
- Discourse flow vs. Process flow
- Activity identification
- Agent / Object identification
- NLP related issues



# Related literature

Paper	Input	Output	Method
Thorne et al. (2013)	Clinical documents	Process fragment	<ul style="list-style-type: none"><li>• NLP techniques</li><li>• Supervised method</li></ul>
Gonçalves et al. (2010)	Narratives of process workers	Process models	<ul style="list-style-type: none"><li>• NLP techniques</li><li>• External domain knowledge</li></ul>
Ghose et al. (2007)	Corporate documentation	Proto-models	<ul style="list-style-type: none"><li>• NLP techniques</li><li>• External domain knowledge</li><li>• Templates matching</li></ul>
Friedrich et al. (2011)	Text of business processes	Process models	<ul style="list-style-type: none"><li>• NLP techniques</li><li>• Unsupervised method</li></ul>

# Our solution



# NLP Parser

- Stanford Parser

(  
All archaeological features were recorded in plan  
(  
using a Leica 1200 series SmartRover GPS  
)  
and surveyed in accordance with CA Technical  
Manual 4 Survey Manual (2012)  
)

# Verb semantics

« The archaeologist directed a survey. »



**VerbNet**

NP.Agent Verb  
NP.Location

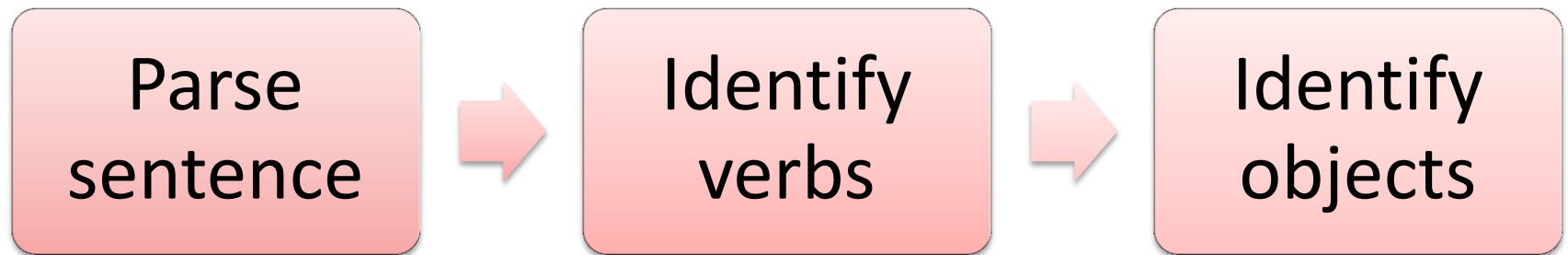
**WordNet**

Somebody Verb  
Something

transitive & active **Verb Knowledge Base**

# ActivityMiner

- VerbNet + WordNet => Verb Knowledge Base





# Example

(ROOT  
  (S  
    (NP (DT All) (JJ archaeological) (NNS **features**))  
    (VP (VBD were)  
      (VP  
        (VP (**VBN recorded**)  
          (PP (IN in)  
            (NP (NN plan))))  
        (S  
          (VP (**VBG using**)  
            (NP (DT a) (NNP **Leica**) (NNP 1200)  
              (NN **series**) (NNP **SmartRover**) (NNP **GPS**))))))  
      (CC and)  
      (VP (**VBN surveyed**)

## Output: **Activities**

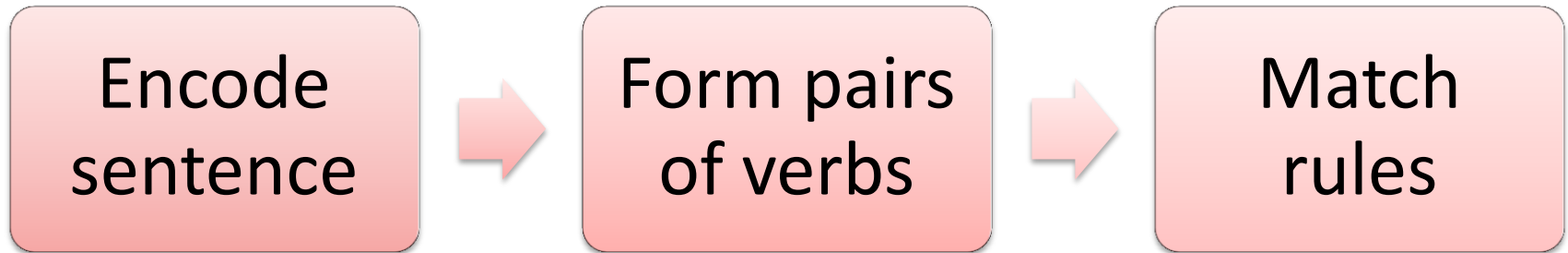
Record feature

Use Leica\_series\_Smartover\_GPS

Survey feature

# ActivityRelationshipMiner

- Rules Matching Mechanism:
  - If ... then / When
  - And / Or



# Example

- “... All archaeological features were recorded in plan using a Leica 1200 series SmartRover GPS and surveyed in accordance with CA Technical Manual 4 Survey Manual (2012)... “

# Example

(S ... 5.VBN ... (S 6.VBG ... ) and 7.VBN ... .)

Pairs:

- 5.VBN ... 6.VBG
- 5.VBN ... and 7.VBN

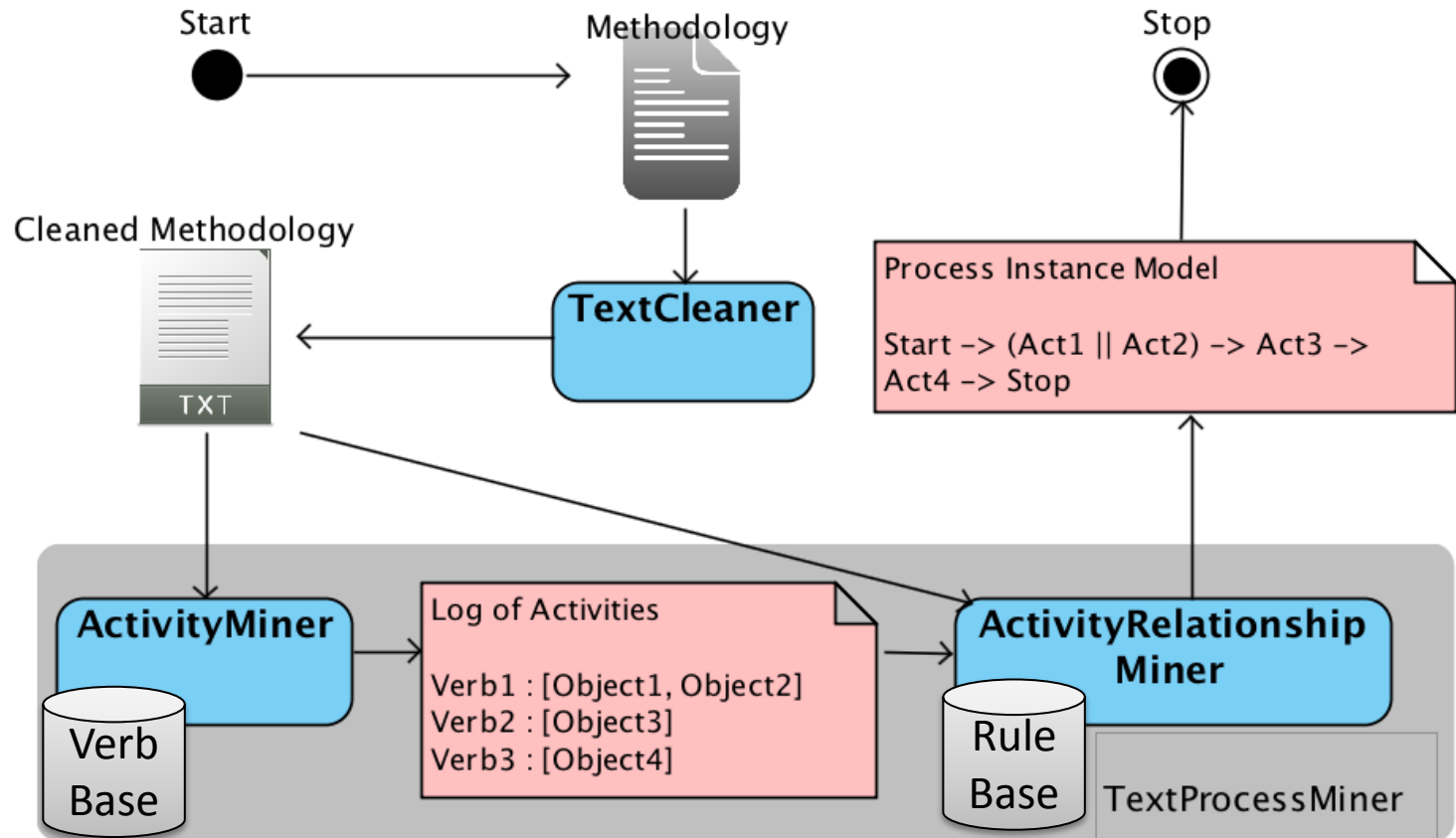
*1.VB/VBD/VBN ... 2.VBG => 1||2*

*1.VB/VBD/VBN/VBG ... and ... 2.VB/VBD/VBN/VBG => 1->2*

Output: **Process Instance Fragment**

(5 || 6) -> 7

# Solution overview





# Validation - Villa-Magna Project



About

Books 10

## Elizabeth Fentress

Research Interests: [Archeology of the longue durée in the Western Mediterranean](#), [History of slavery in the Mediterranean](#),



# Validation

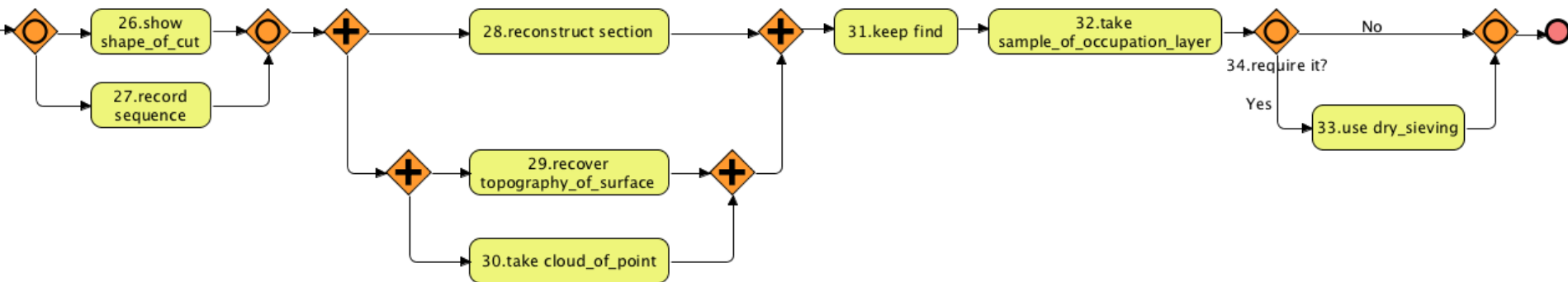
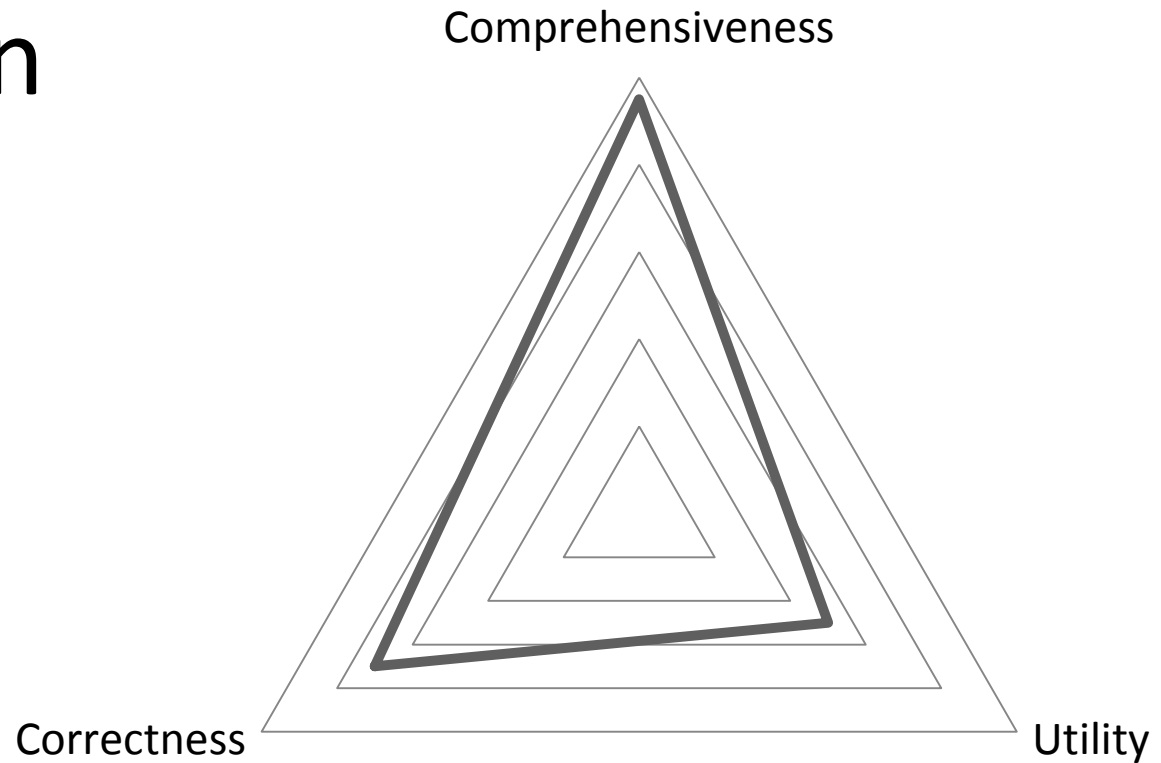
- Activities:
  - Correctness, Completeness
- Process instance model:
  - Correctness, Comprehensiveness, Utility

# Validation

- 34 activities found:
  - 4 false positives
  - 3 wrong names
  - 1 missing activity

1. carry\_out excavation
2. use technique\_of\_context\_area\_excavation
3. codify technique\_of\_context\_area\_excavation
4. strip topsoil
5. trowel deposit
6. define uppermost\_layer
7. define unit
8. plan uppermost\_layer
9. plan unit

# Validation



# Lessons learnt

+

- High precision of ActivityMiner
- Satisfactory process instance model
- Comprehensive process instance model

-

- Process instance model only
- Sentence ordering
- Sparse conditionality

# Takeaways

- Problem
  - Frequent, recurrent and difficult
- Solution
  - Unsupervised, NLP-based, no domain knowledge
  - Focus on verb semantics
- Further
  - how to create process models?