# STORY
# Spatio-Temporal Object Repository

## Massimiliano Albanese
## Università di Napoli "Federico II"

Scuola per Dottorandi in Ingegneria dell'Informazione

Napoli, 16-20 Febbraio 2004

# Introduction

- STORY is
  - a Spatio-Temporal Object Repository
  - a system to dynamically generate and deliver "stories" about people, events and places
    - According to the Oxford English Dictionary a **story** is "*A narrative, true or presumed to be true, relating to important events and celebrated persons of a more or less remote past; a historical relation or anecdote.*"
    - In the context of computing, narratives can be rendered as interactive multimedia presentations

# STORY Project

- Joint research project
  - University of Maryland, College Park, USA
    - V.S. Subrahmanian
    - Marat Fayzullin
  - Università di Napoli "Federico II"
    - Antonio Picariello
    - Massimiliano Albanese
    - Carmine Cesarano

# A possible scenario
## Pompeii archaeological site

- Pompeii is a spectacular archaeological site.
- Visitor experience can be greatly improved by:
  - Automatically notifying visitors of interesting phenomena without posting extra signs
  - Allowing visitors to explore the stories of various monuments, paintings, sculptures, etc. in Pompeii.
  - Allowing visitors to explore the stories of the characters, events and places depicted in these monuments, paintings, sculptures, etc.
- Visitors interests vary – so information about exhibits must adapt in real time to their interests to enhance the experience of the visitor.
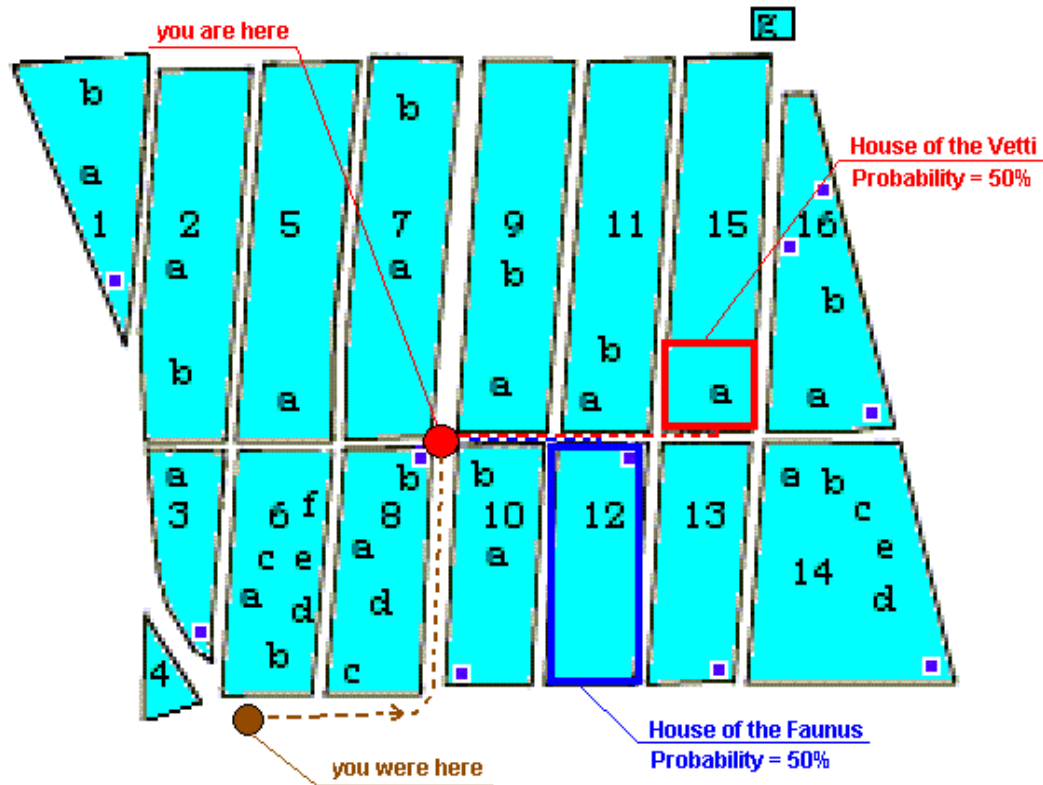
# Pompeii Visitors



Visitor arrives at ticket counter and buys ticket.
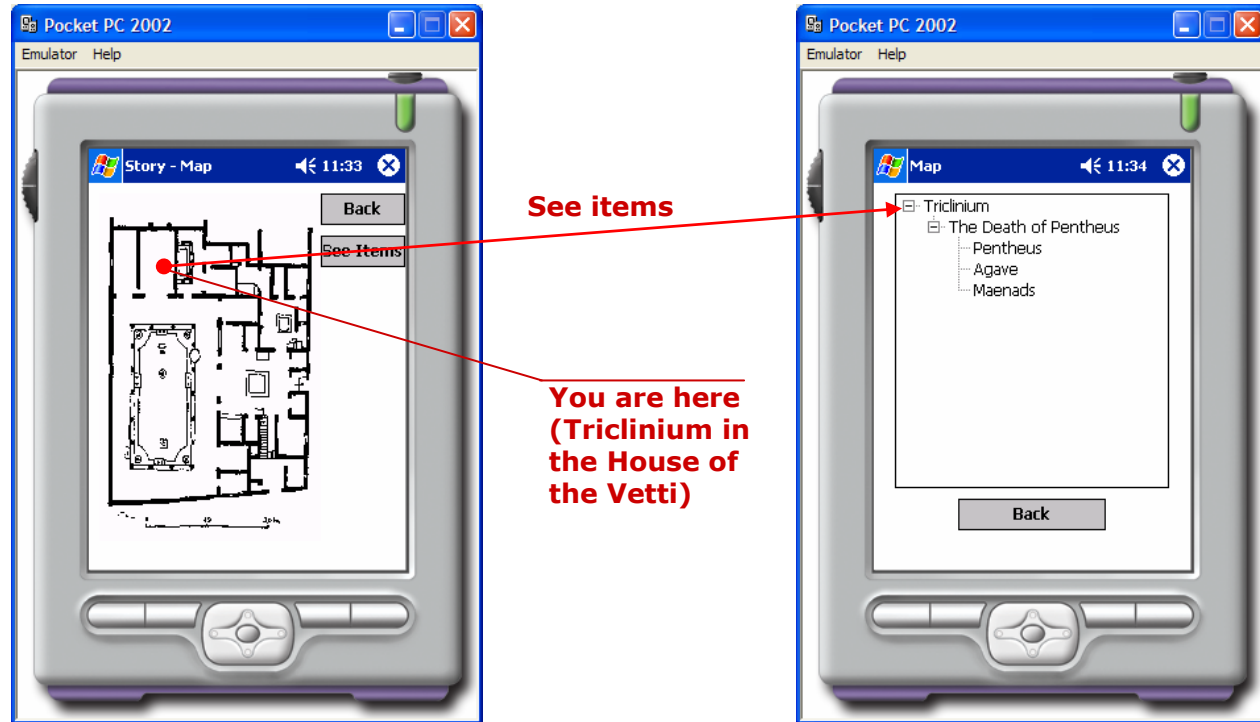
# Pompeii Visitors



Ticket agent asks if they would like to use the story facility and if they would like to use their cell phone and/ or PDA to get stories of interest to them.
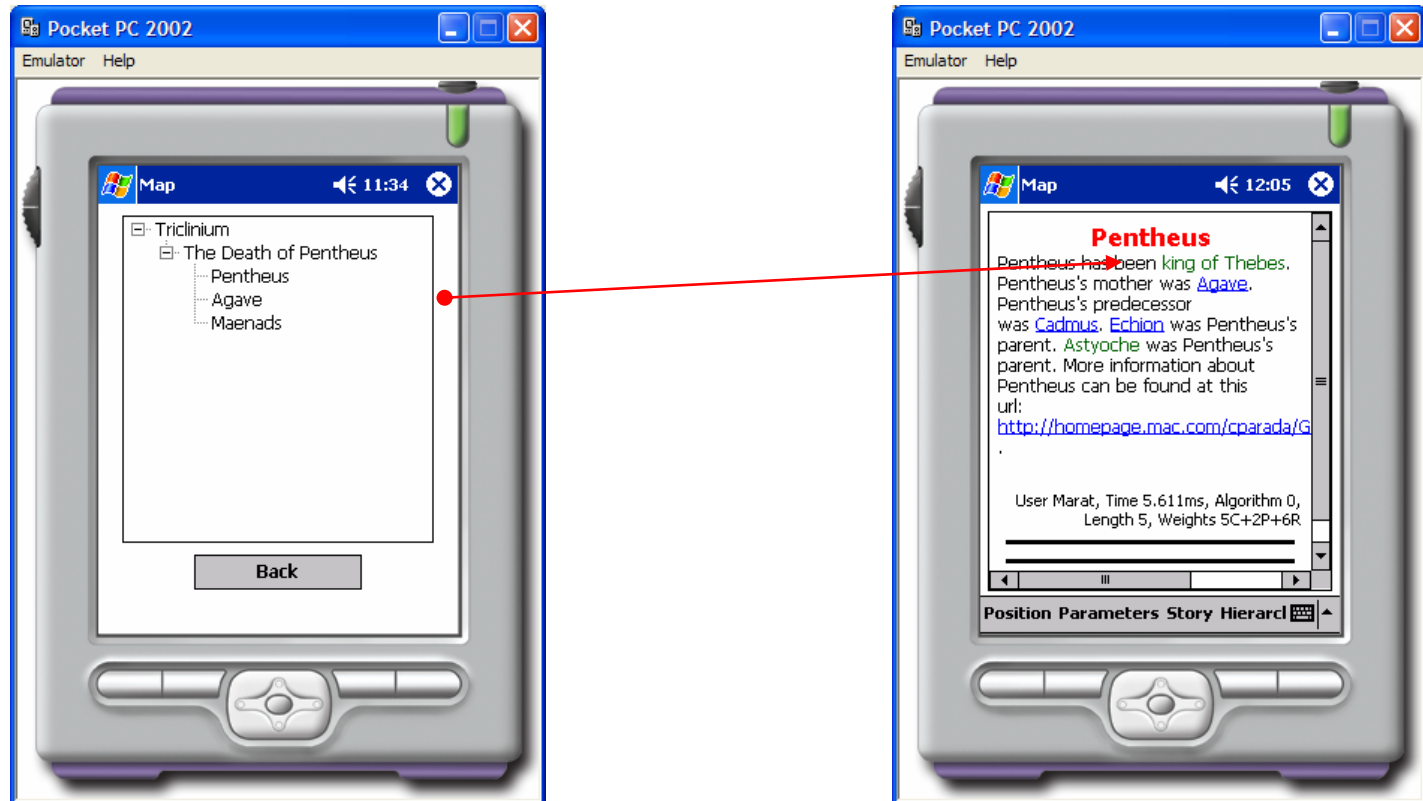
# Pompeii Visitors



As visitor walks through Pompeii, STORY identifies where he is and predicts where he might go in the future (probabilistically). Ex. if he is at location L, it might predict that he will go to the House of the Vetti.

# Pompeii Visitors



**See items**

**You are here (Triclinium in the House of the Vetti)**

Based on this prediction of where he might go in future, it identifies potential stories he might be interested in and downloads parts of these stories to his PDA/cell. E.g. It might download stories about Pentheus.

# Pompeii Visitors



The visitor chooses which story he is interested in. STORY dynamically generates the story and delivers it to the user's PDA/cell phone, e.g. user might choose story of Pentheus.

# Pompeii Visitors
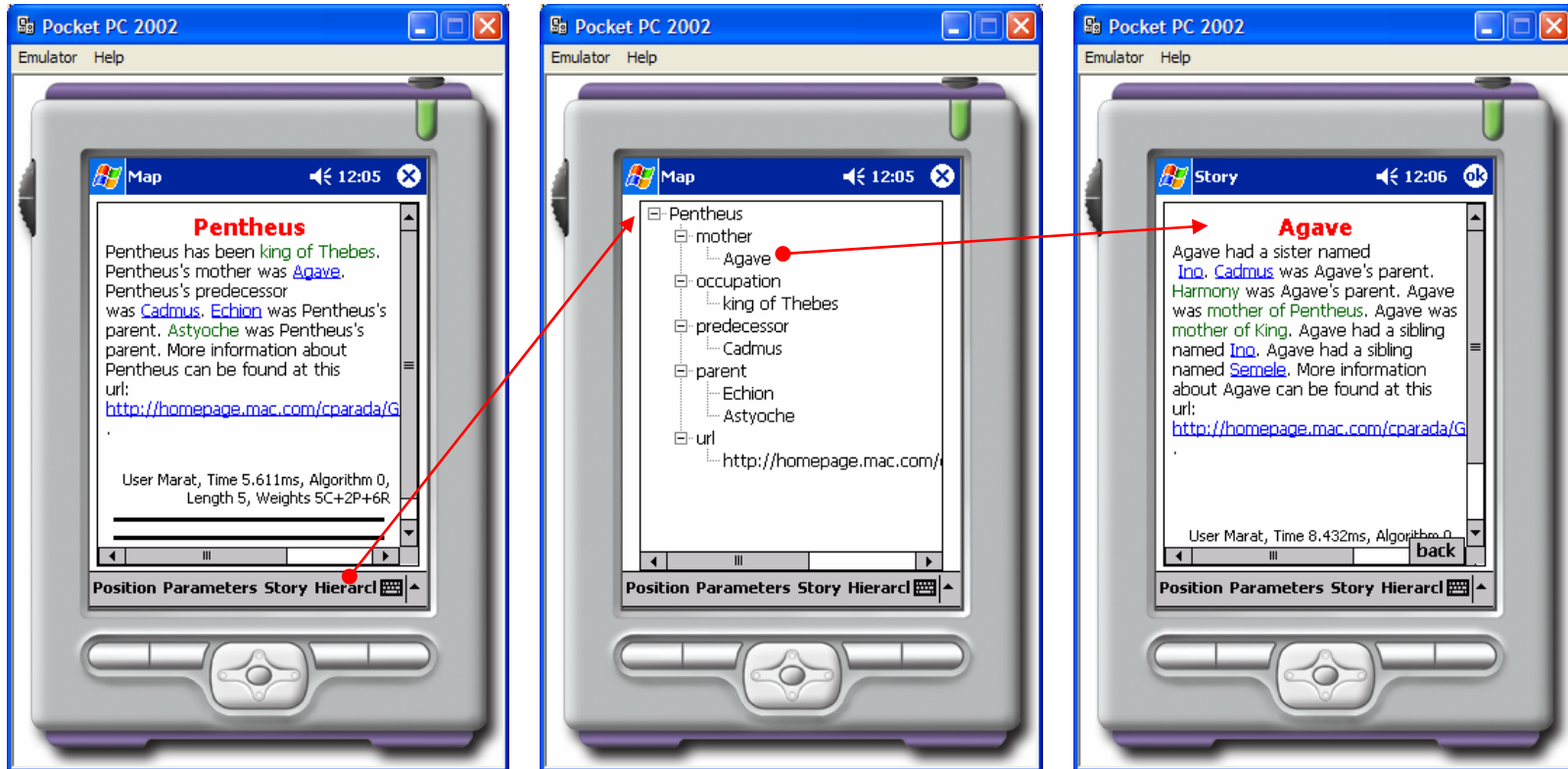


The user can choose to explore the story in greater detail (e.g. if he is seeing the story of Pentheus, he can also explore the story of Agave).

# What is a story?

- Set of facts about an entity (person, event, place, monument, painting, etc.).
- Fact = attribute + value
- EX: story of Pentheus
  - occupation: King of Thebes
  - mother: Agave
  - father: Echion
  - predecessor: Cadmus
  - and so on.
- There may be a LOT of facts known about Pentheus. Not all can be presented.

# Story properties

- Stories must
  - include important (high priority) facts
  - be continuous
  - not be repetitive
  - resolve conflicts
- Stories must adapt to visitor interests and dynamically adjust to available bandwidth and output device
  - how we present a story on a cell phone is different from how we might present it on a PocketPC.

# Automatic vs manual story creation

- Automatic and dynamic generation of stories has some advantages versus the manual creation of stories
    - Hard to manually predict all the stories visitors may want to pursue
    - Hard to manually predict all the attributes that different visitors may be interested in
    - Hard to manually predict user navigation patterns *a priori*

# STORY System

- STORY is a system for
  - <u>extracting</u> story content from multiple distributed data sources (databases, web pages, digitized historical documents, maps, etc.)
  - <u>creating</u> a succinct story based on the above content that adapts to user preferences and interests in real time and
  - <u>delivering</u> these stories to users across both wireless, wired, and cellular networks and multiple output devices.

# STORY Components

- **Story configuration module**
  - what is important to a story domain?
- **Story content extraction engine**
  - what facts constitute a story?
- **Story database**
  - where and how should the story content be stored to avoid future extraction?
- **Story creation engine**
  - how can the above facts be woven into a succinct and attention grabbing story?
- **Story delivery engine**
  - how to deliver the story over PDAs/cell phones/Internet?
- **Story prediction engine**
  - what stories should we download before the visitor reaches the location?
- **Story mapping engine**
  - how is the site laid out?

# My contribution to the project

- Story configuration module
  - A set of tools to setup the System for a particular domain (an archaeological site, a museum, a battle field)
    - definition of the spatial features of the system (maps, localization, etc.)
    - definition of an attributes ontology (what is important to a story domain?)
    - definition of templates to render the facts in a textual format
- Story content extraction engine
  - Extraction of facts from
    - heterogeneous distributed databases (a consolidated topic)
    - web pages (the major challenge)

# Story content extraction engine
## Extraction of facts from web pages

1. Extraction of text from web pages
2. How to extract facts (attribute + value) about entities of interest from unstructured text?
   - A knowledge base is needed in order for any computing system to understand a piece of text and infer new knowledge from it
     - WordNet is a well-consolidated lexical reference system and it has been adopted as the lexical reference of our system

# Story content extraction engine
## Extraction of facts from web pages

- An algorithm for extraction of facts from text
  - Lexical analysis by means of WordNet
    - Which is the role of each word in the text?
  - Syntactic analysis
    - Text tokenization
    - Named entities recognition
    - Part of speech disambiguation
    - Pronouns resolution

# Story content extraction engine
## Extraction of facts from web pages

- **Text rewriting**
  - Based on the result of the previous step, text can be rewritten in a cleaner form
- **Semantic analysis**
  - Word sense disambiguation
  - Application of rules for text pattern recognition and information extracton
  - Rule are in the form <head> → <tail>
    - if the <head> is satisfied a fact about an entity is determined according to the <tail>
  - The system can learn rules starting from a set of examples

# Further works

- Several aspects require further investigation
  - More complex learning schemes for the extraction engine
  - Management of event timelines
  - Management of conflicting information and different points of view
  - Prediction of user behavior