

Tutorial:

Procesamiento de Lenguaje Natural en Ingeniería de Requisitos: *Contribuciones Potenciales y Desafíos de Investigación*

CIBSE 2015 – Lima, Peru

Dr Ricardo Gacitúa
ricardo.gacitua@ceisufro.cl

Centro de Estudios en Ingeniería de Software
Depto. Cs. Computación e Informática
Universidad de La Frontera
Temuco, Chile

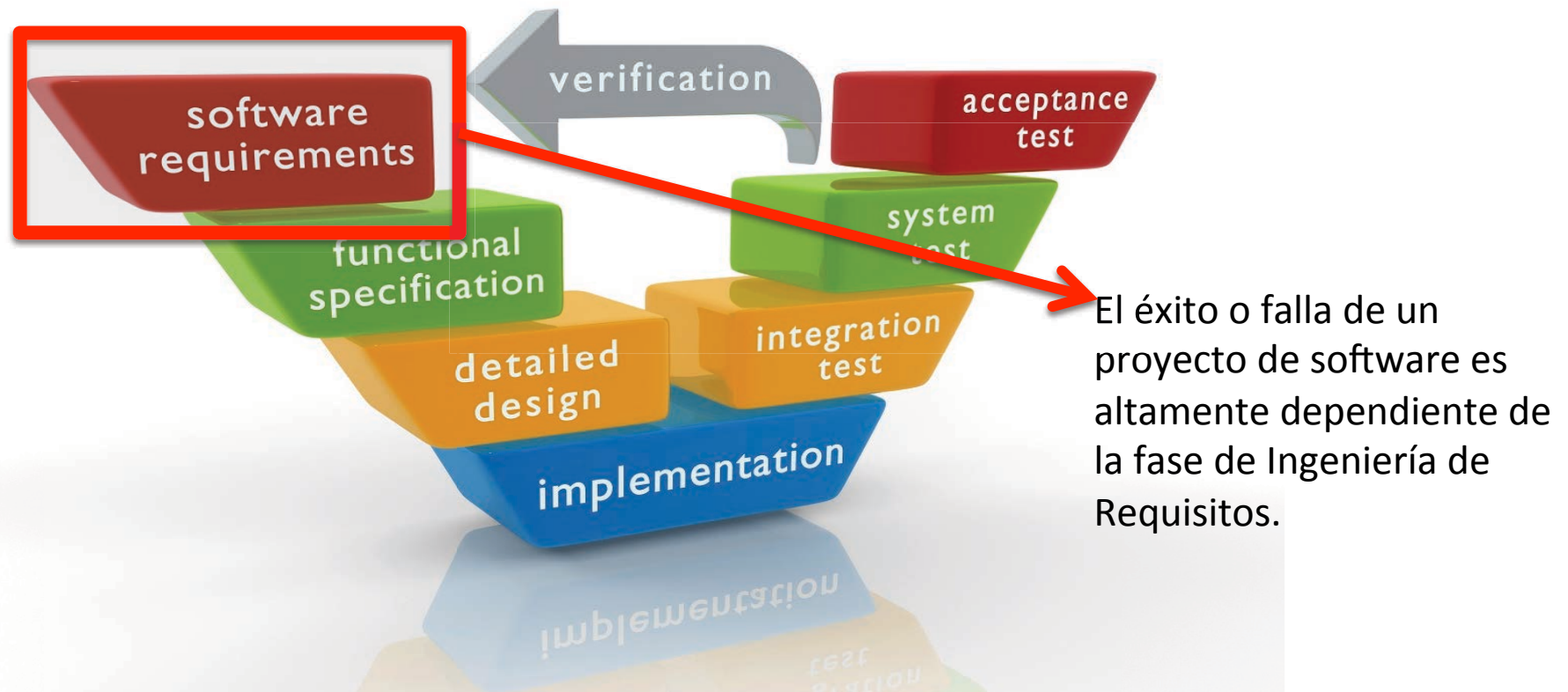
Indice

- **Ingeniería de Requerimientos (RE)**
- **Procesamiento de Lenguaje Natural (NLP)**
- **Avances en la relación entre RE & NLP**
- **Contribuciones Potenciales**
- **Desafíos de Investigación**

Después de este tutorial, podrás :

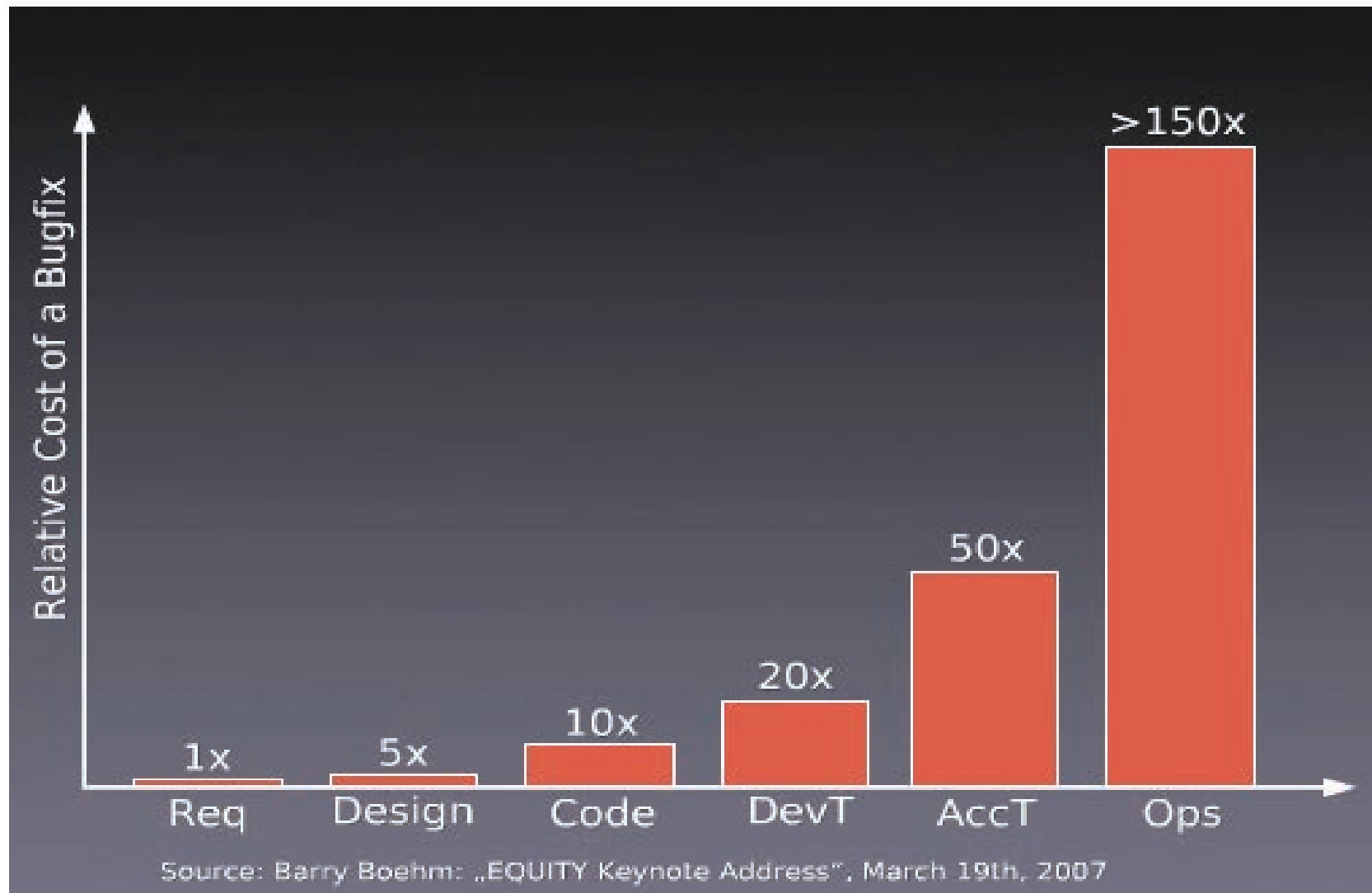
- **Definir RE y su describir su importancia.**
- **Definir NLP y sus áreas de acción.**
- **Conocer el estado del arte en la relación entre NLP y RE.**
- **Describir las contribuciones potenciales de NLP para RE.**
- **Listar los principales desafíos para la investigación en RE.**

Software Engineering Process

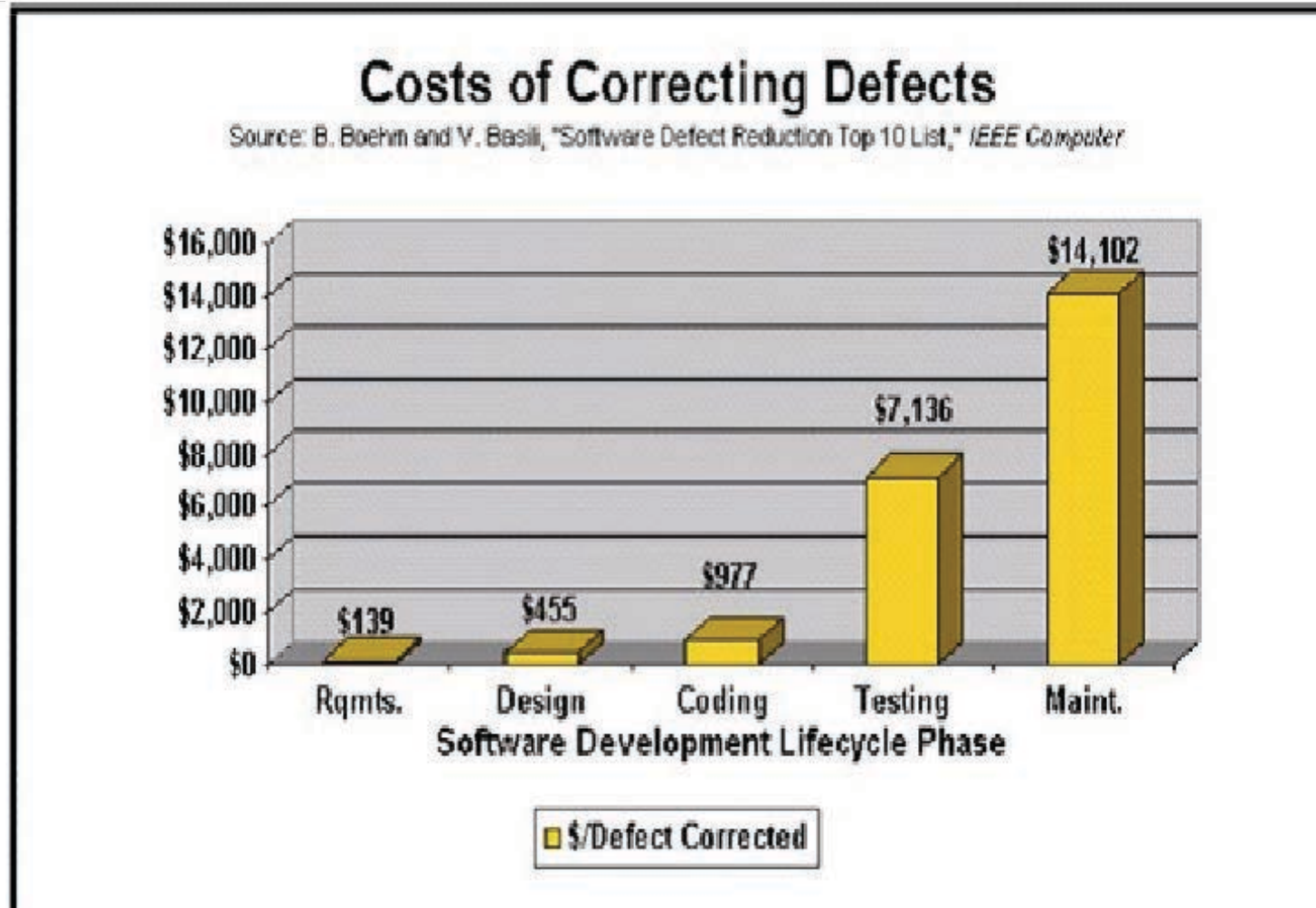


- Es una división del trabajo de desarrollo de software en distintas fases (o etapas) que tienen actividades las cuales persiguen una mejor planificación y administración.

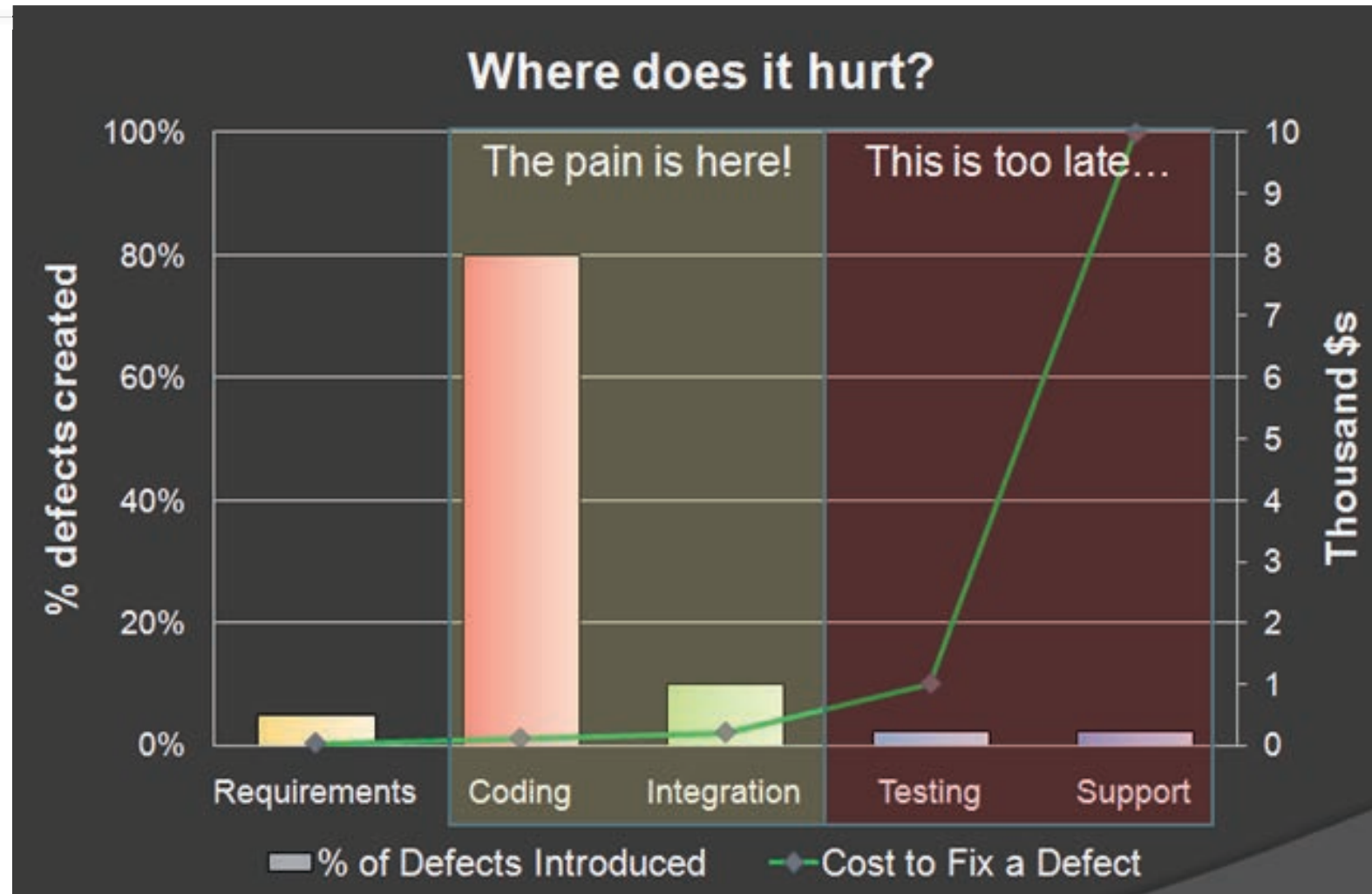
Increasing Cost of Defects



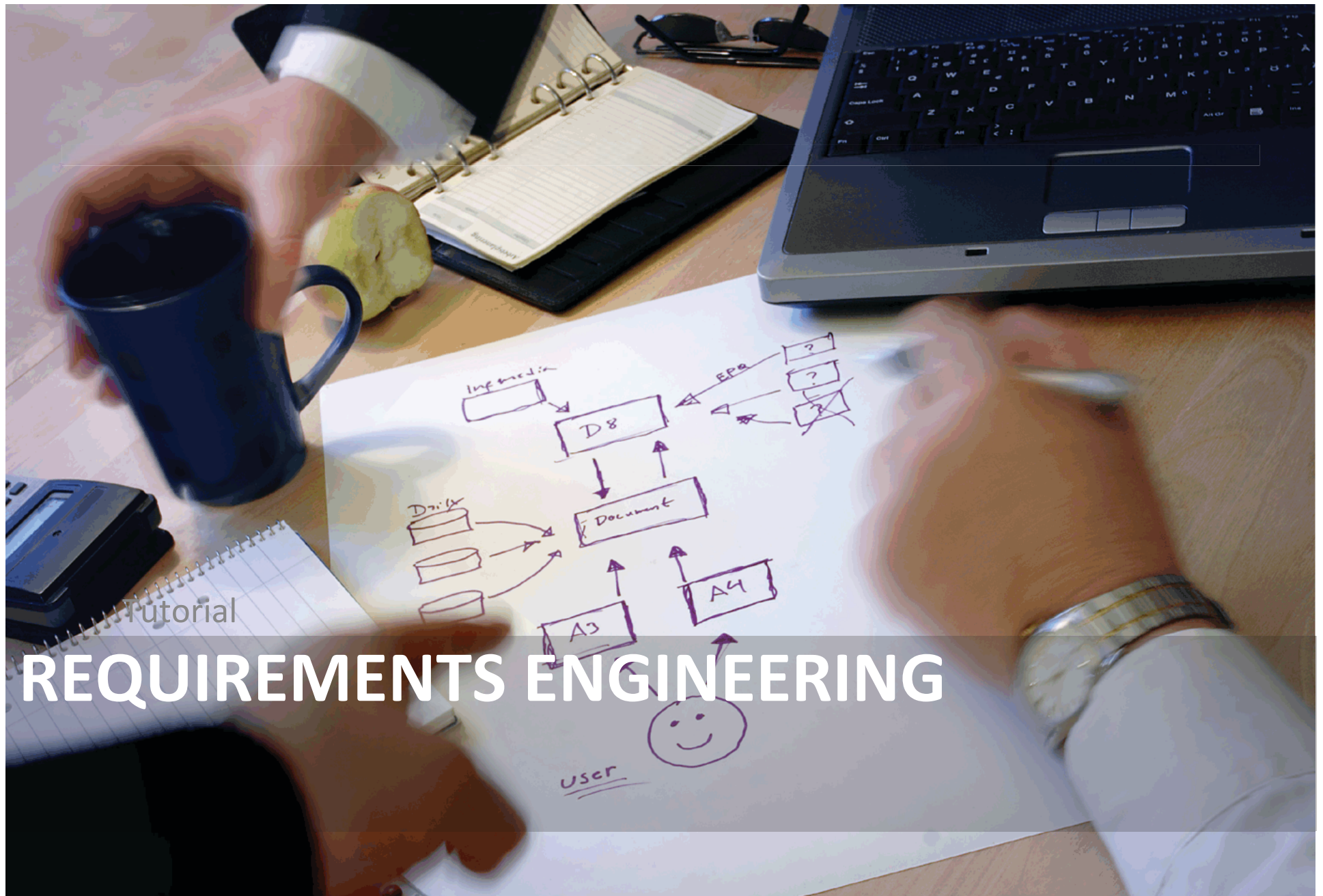
Cost of Correcting Defects



Defects Introduced



Graphs taken from: *Benefit From Unit Testing in The real world.*



Tutorial

REQUIREMENTS ENGINEERING

Requirements Engineering



DEFINITION Requirements engineering is the branch of software engineering concerned with the **real-world goals** for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to **precise specifications** of software behavior, and to their **evolution over time** and across software families.

Zave, L. (1997). Classification of Research Efforts in Requirements Engineering. ACM Computing Surveys, 29(4): 315-321.

“Real-world goals” represent:

- the ‘why’ as well as
- the ‘what’ of a system.

“Precise specifications”
provide the basis for:

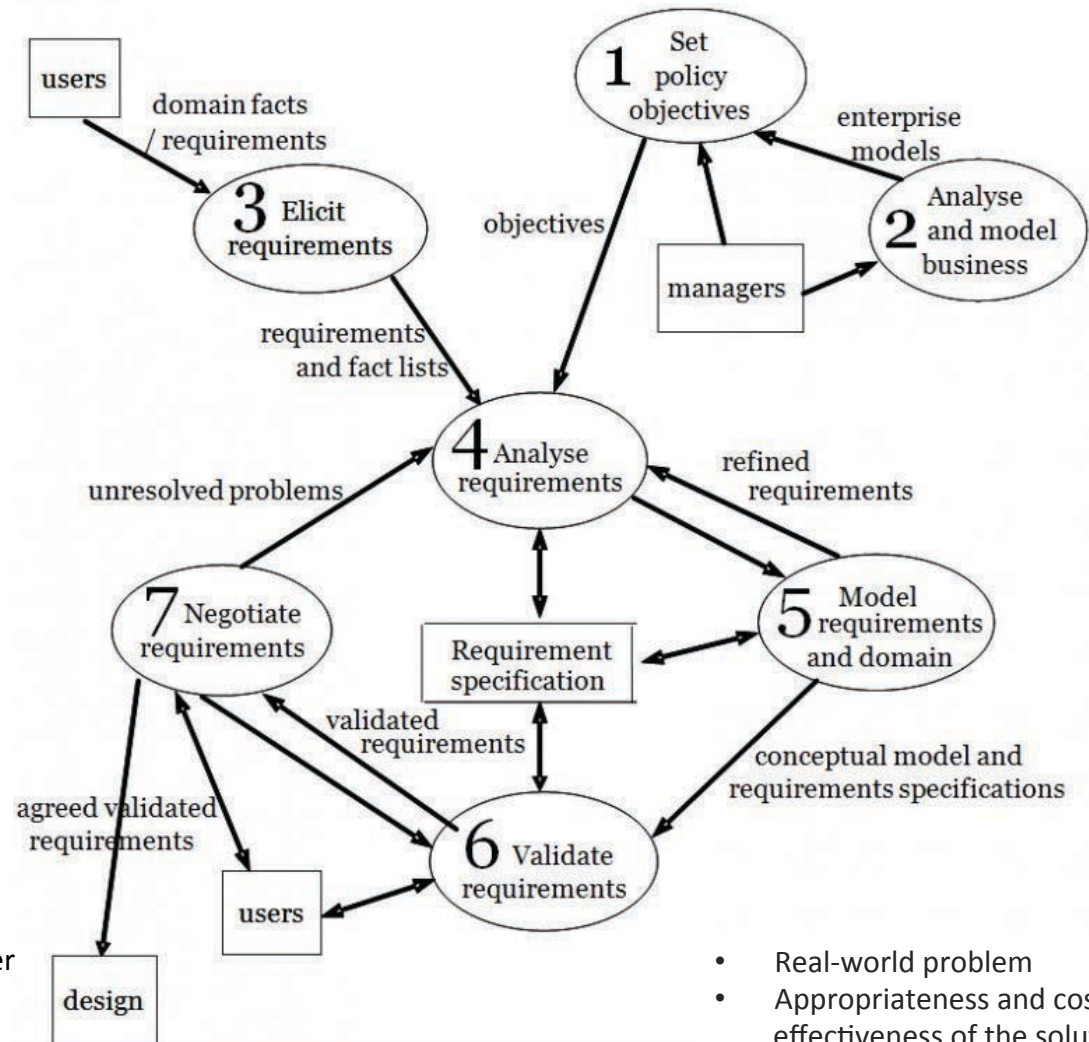
- Analysing requirements,
- Validating that they are indeed what stakeholders want,
- Defining what designers have to build, and
- Verifying that they have done so correctly upon delivery.

“Evolution over time and across software families.”

- a changing world and the need to reuse partial specifications.

Requirements Engineering Process

- The use of the term *engineering* in RE serves as a reminder that RE is an important part of an engineering process.



Sutcliffe, Alistair G. (2014): Requirements Engineering. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark:

- Real-world problem
- Appropriateness and cost-effectiveness of the solution

RE draws on the cognitive and social sciences

This provides an understanding of the difficulties people may have in describing their needs

Cognitive psychology

This provides a methodological approach to observing human activities

Anthropology

This provides an understanding of the political and cultural changes caused by computerisation

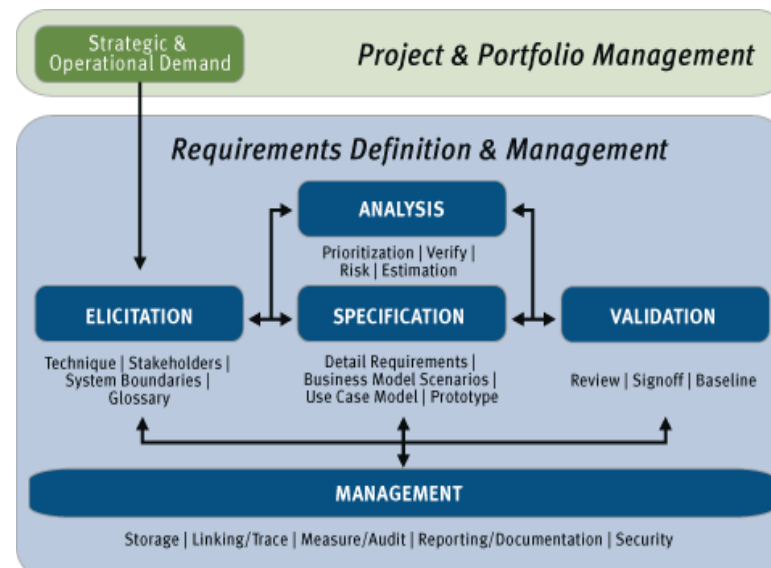
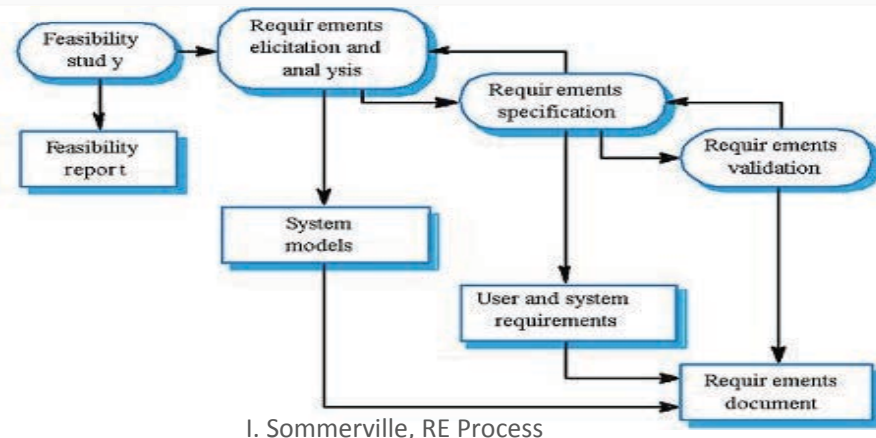
Sociology

This is important because RE is largely about communication

Linguistics

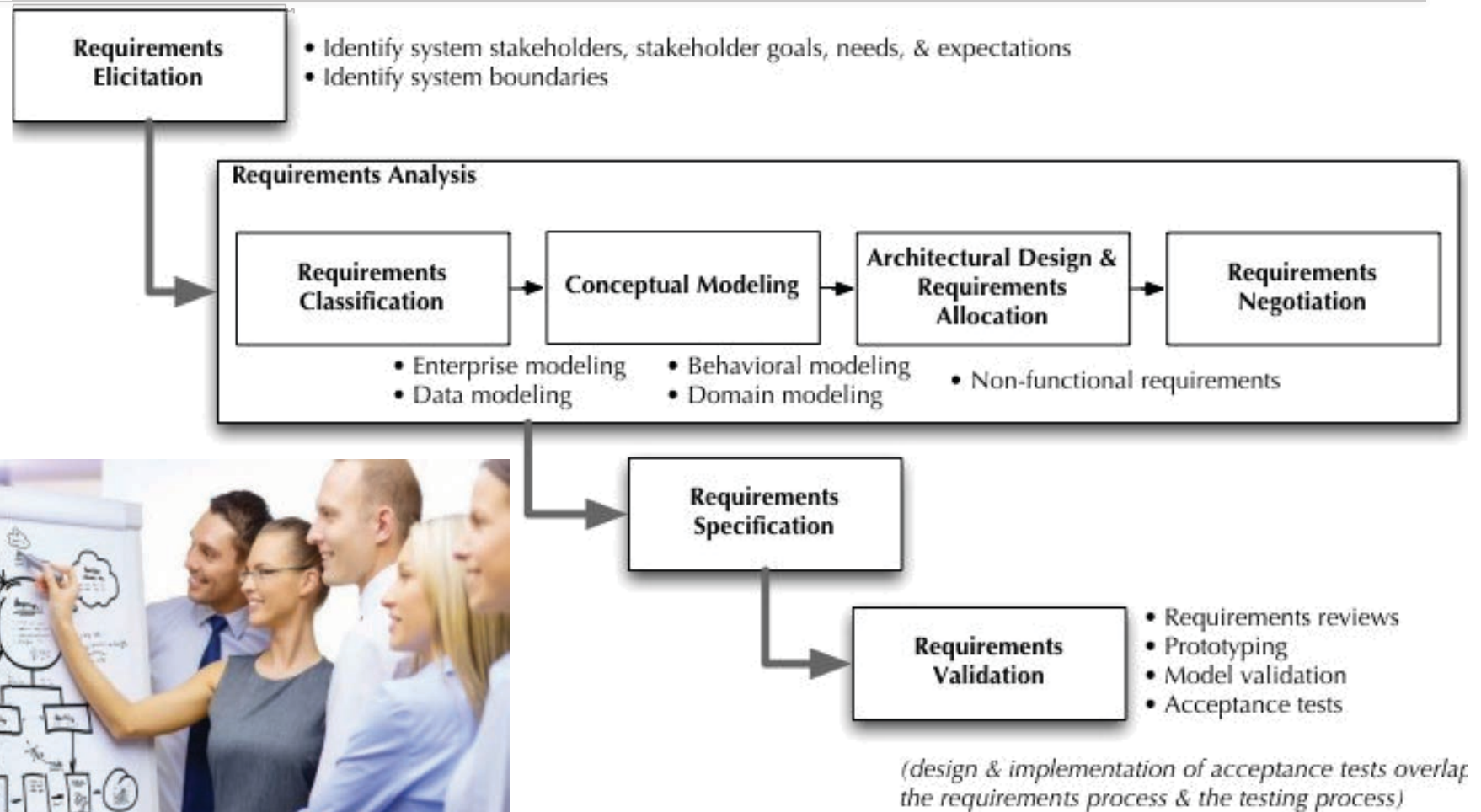
The context in which RE takes place is usually a human activity system, and the problem owners are people.

Requirements Engineering Process



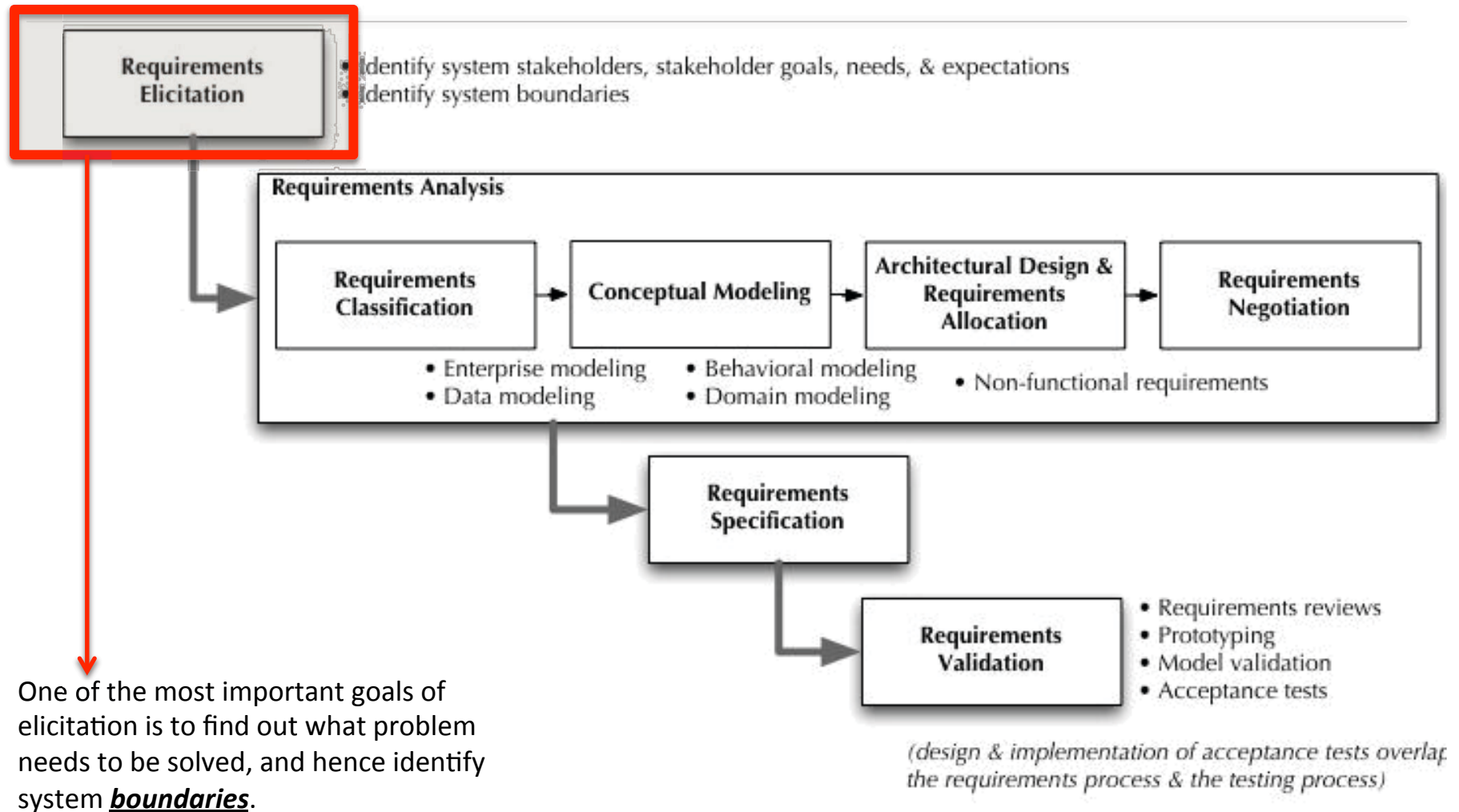
Micro Focus CaliberRM™ - Enterprise Software Requirements Management System

Requirements Engineering Process



http://www.phoenix.tc-ieee.org/015_Methodology_and_Systems_Management/SwE_workflow.html

RE Process -> Elicitation

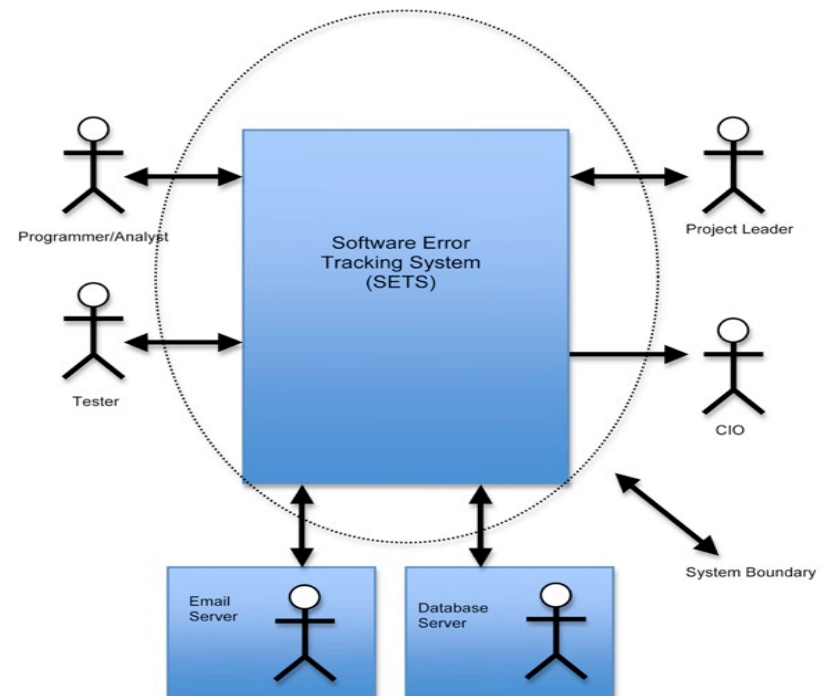
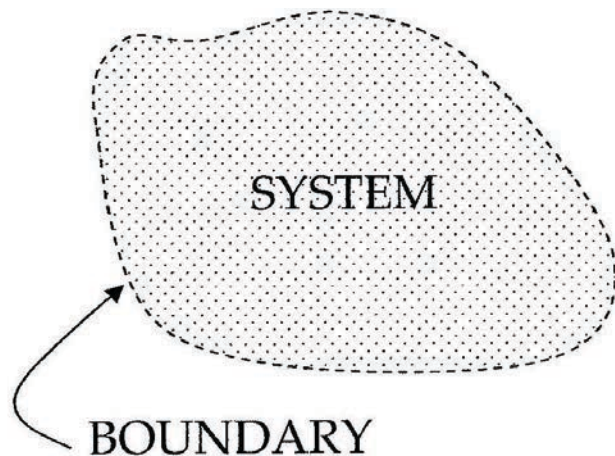


Requirements Elicitation Problems

- Problems of scope
 - the boundary of the system is ill-defined
 - unnecessary design information may be given
- Problems of understanding
 - users have incomplete understanding of their needs
 - users have poor understanding of computer capabilities and limitations
 - analysts have poor knowledge of problem domain
 - user and analyst speak different languages
 - ease of omitting “obvious” information
 - conflicting views of different users
 - requirements are often vague and untestable, e.g., “user friendly” and “robust”
- Problems of volatility
 - requirements evolve over time

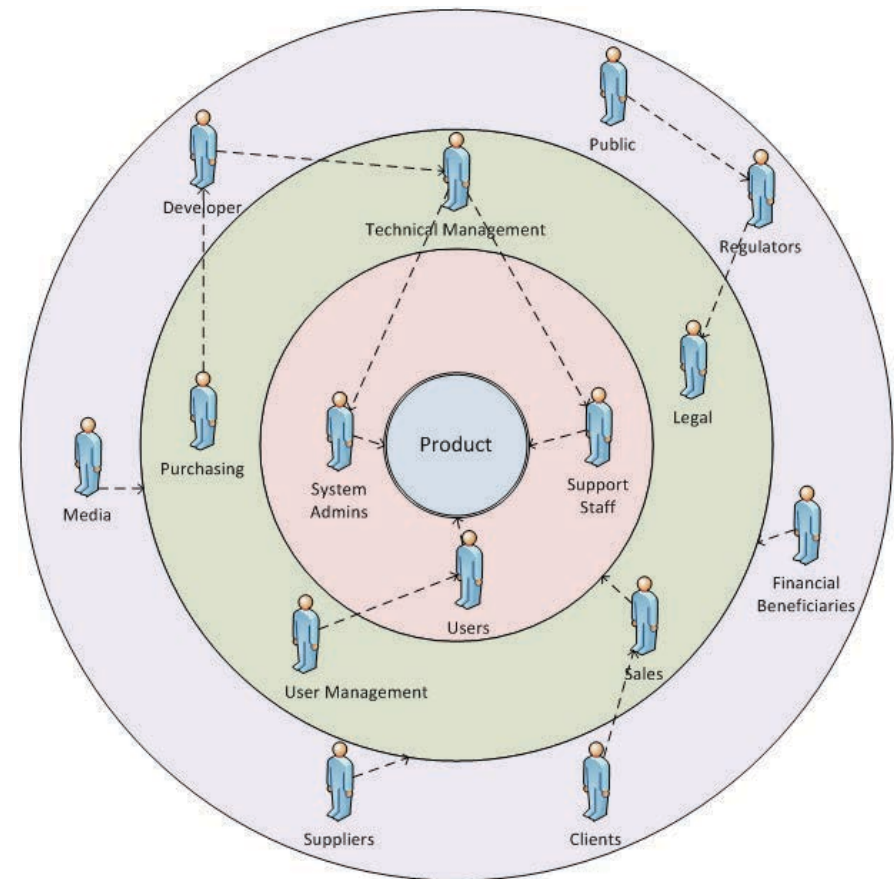
RE Process -> Elicitation

- **A) BOUNDARIES** - Identifying and agreeing a system's boundaries affects all subsequent elicitation efforts.
 - The identification of stakeholders and user classes, of goals and tasks, and of scenarios and use cases all depend on how the boundaries are chosen.



RE Process -> Elicitation

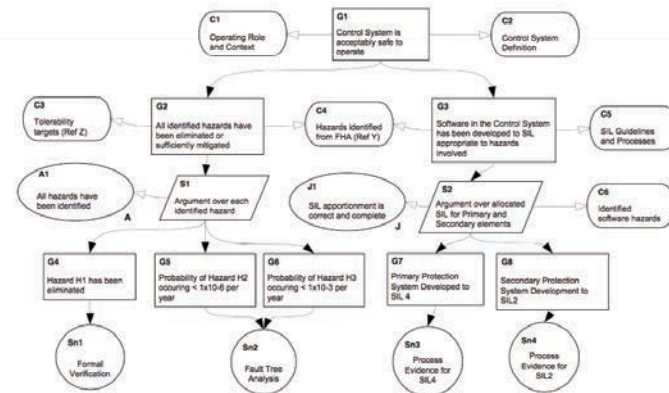
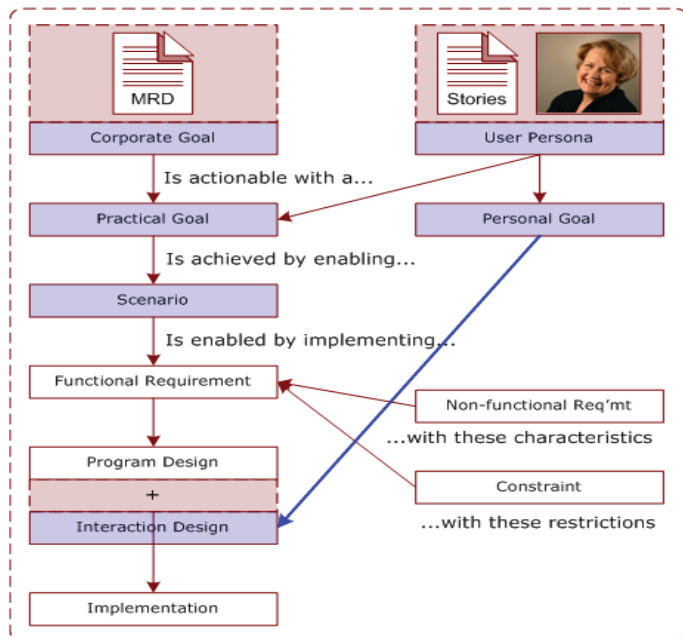
- **B) IDENTIFYING STAKEHOLDERS-** individuals or organisations who stand to gain or lose from the success or failure of a system – is also critical.
 - Stakeholders include:
 - customers or clients (who pay for the system),
 - developers (who design, construct and maintain the system), and
 - users (who interact with the system to get their work done).



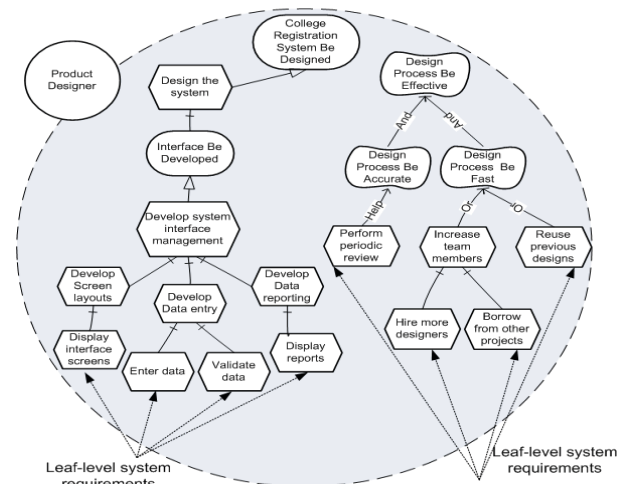
Alexander, I., "Stakeholders: who is your system for?," *Computing & Control Engineering Journal* , vol.14, no.2, pp.22,26, April/May 2003

RE Process -> Elicitation

- C) GOALS - denote the objectives a system must meet.



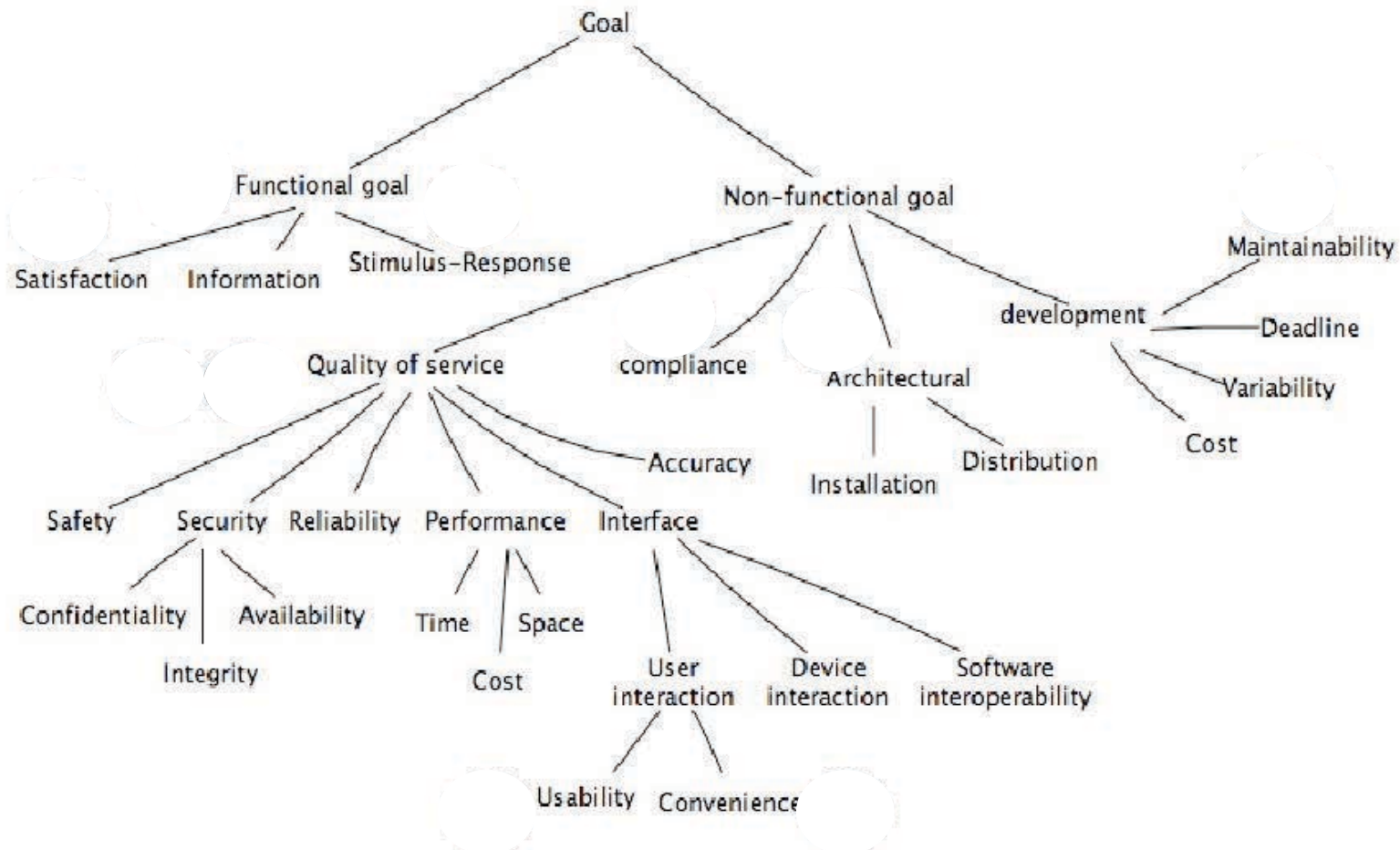
Example of Goal Structuring Notation(GNS)



Example i* notation

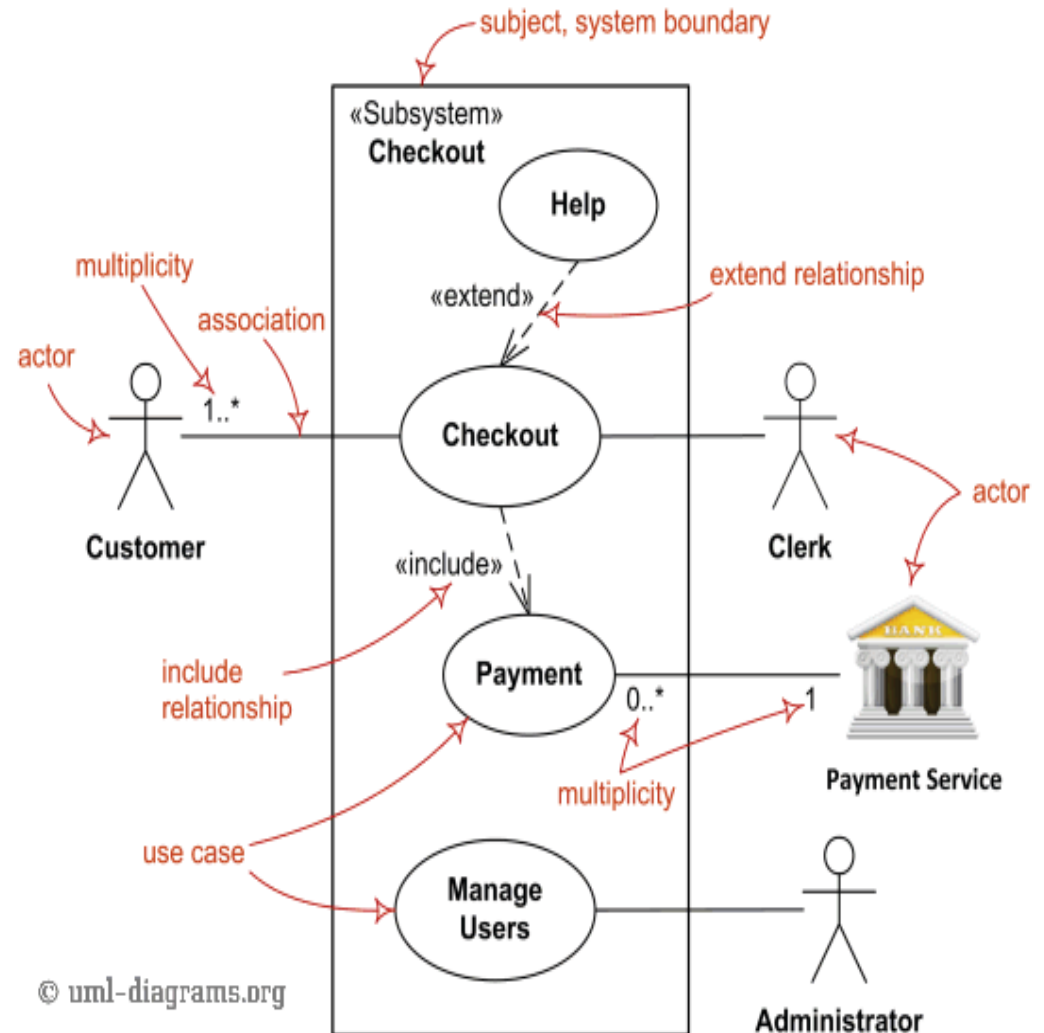
Eliciting goals focuses the requirements engineer on the problem domain and the needs of the stakeholders, rather than on possible solutions to those problems.

RE Process -> Elicitation- Goal categorization

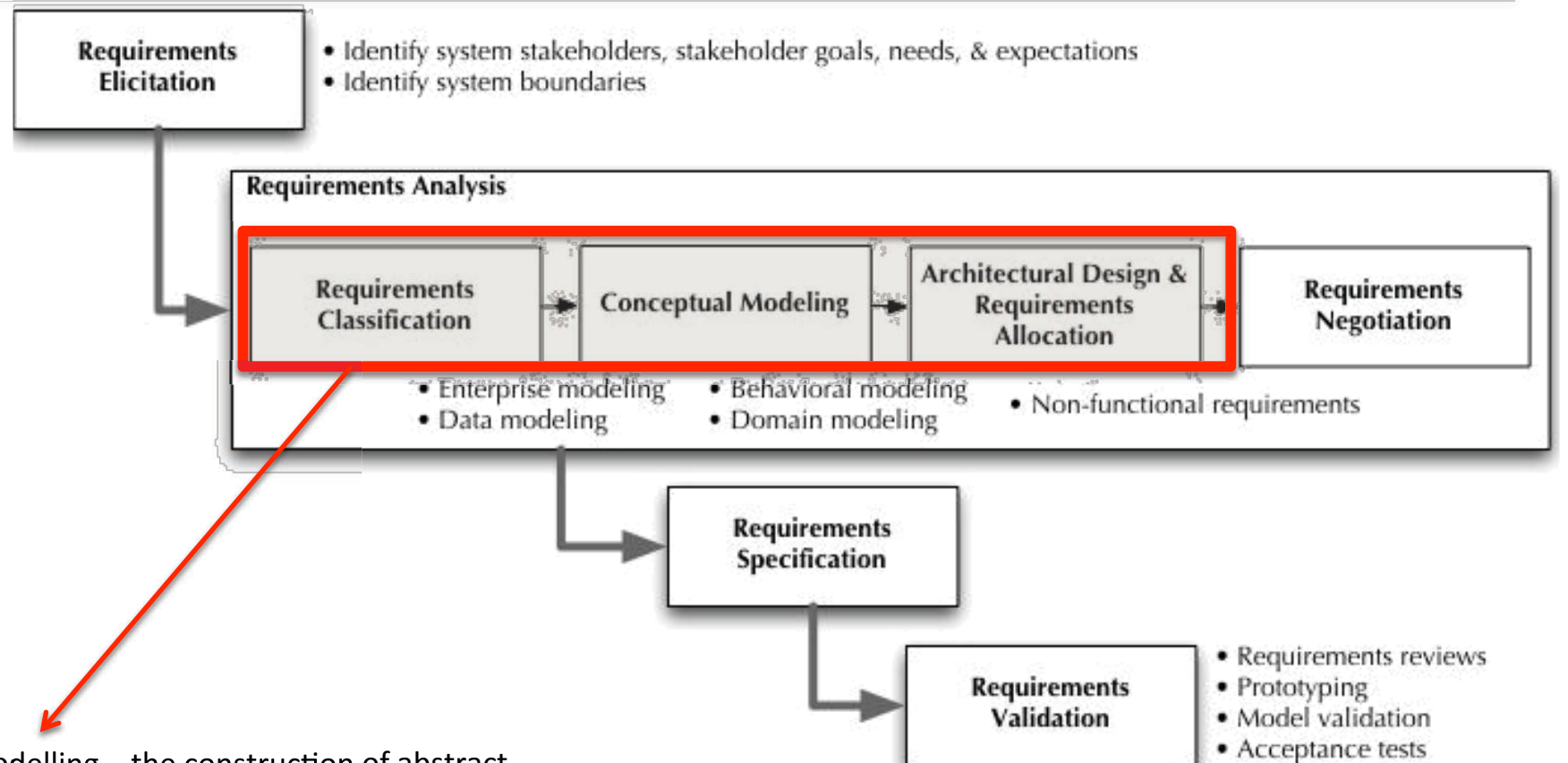


RE Process -> Elicitation - Use Cases to represent user tasks

- It is often the case that users find it difficult to articulate their requirements.
- A requirements engineer can resort to eliciting information about the *tasks* users currently perform and those that they might want to perform.



RE -> Modelling and Analysing Requirements

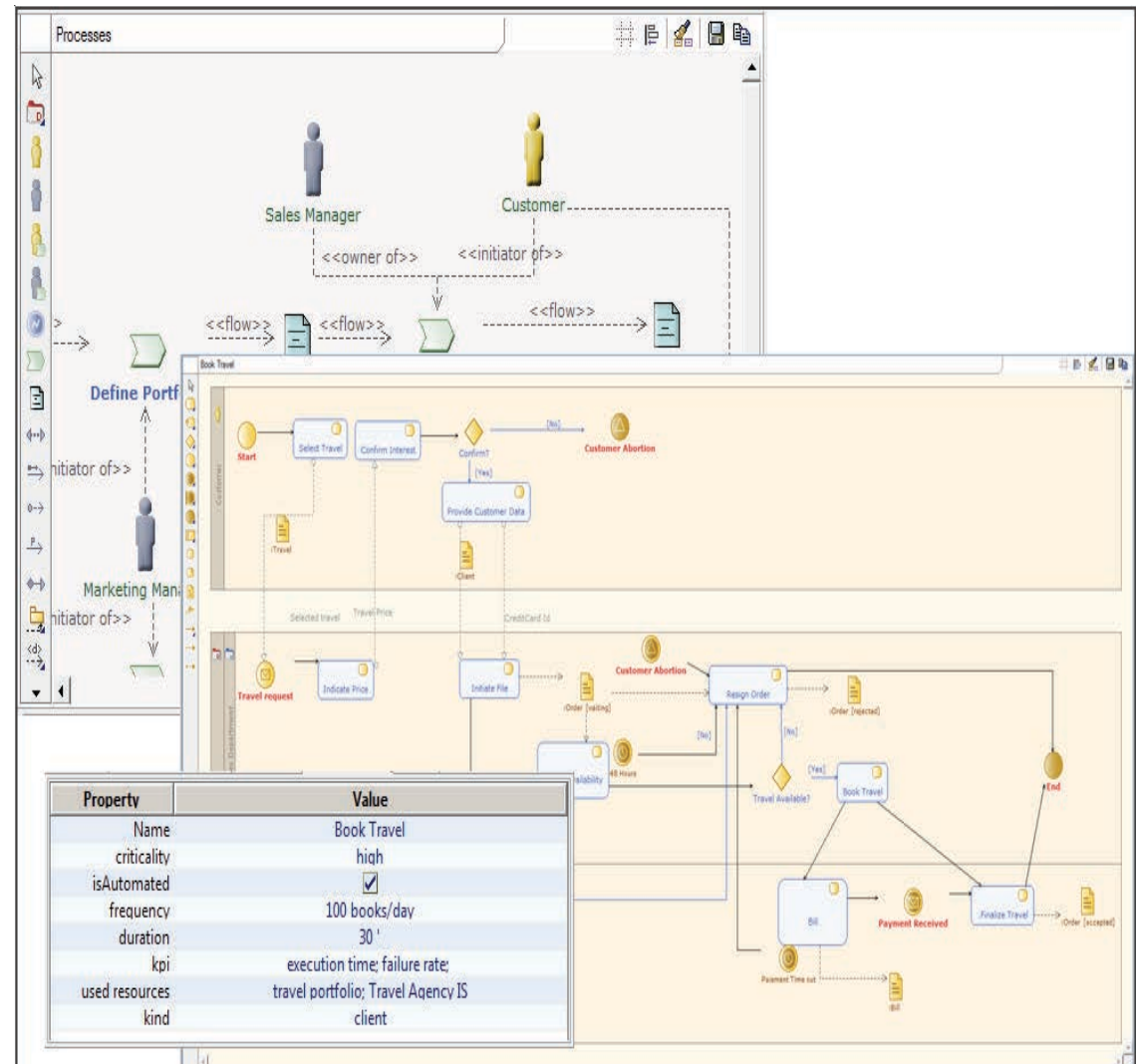


Modelling – the construction of abstract descriptions that are amenable to interpretation – is a fundamental activity in RE.

(design & implementation of acceptance tests overlap the requirements process & the testing process)

RE -> Modelling and Analysing Requirements -

- A) **Enterprise Modelling** - Enterprise modelling and analysis deals with:
 - understanding an organisation's structure;
 - the business rules that affect its operation;
 - the goals, tasks and responsibilities of its constituent members; and
 - the data that it needs, generates and manipulates.

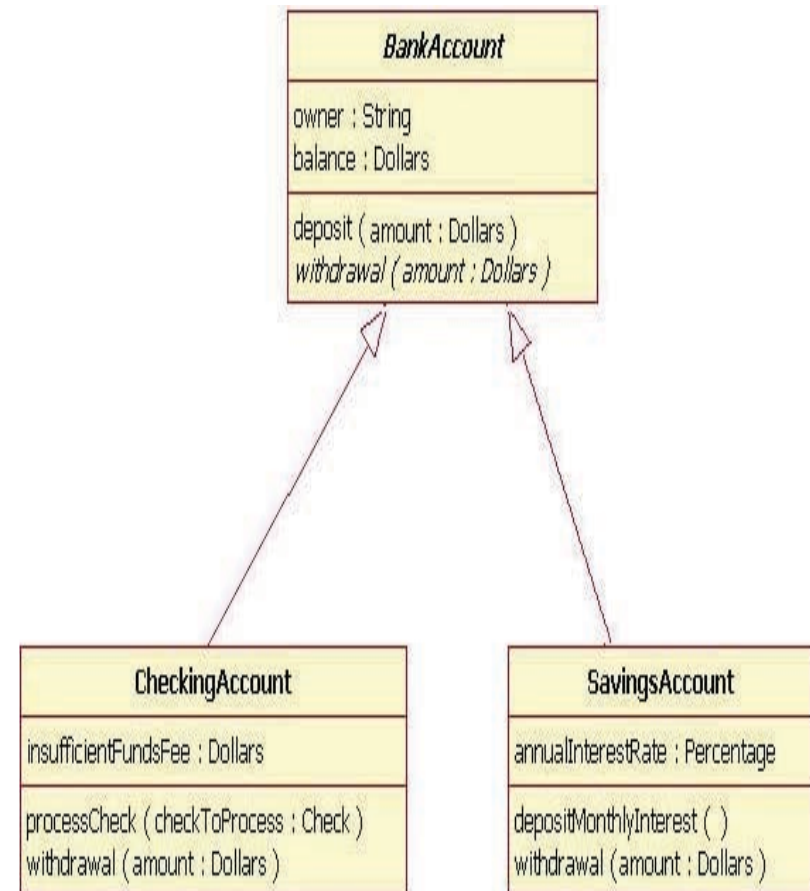


RE -> Modelling and Analysing Requirements -

- B) **Data Modelling** -

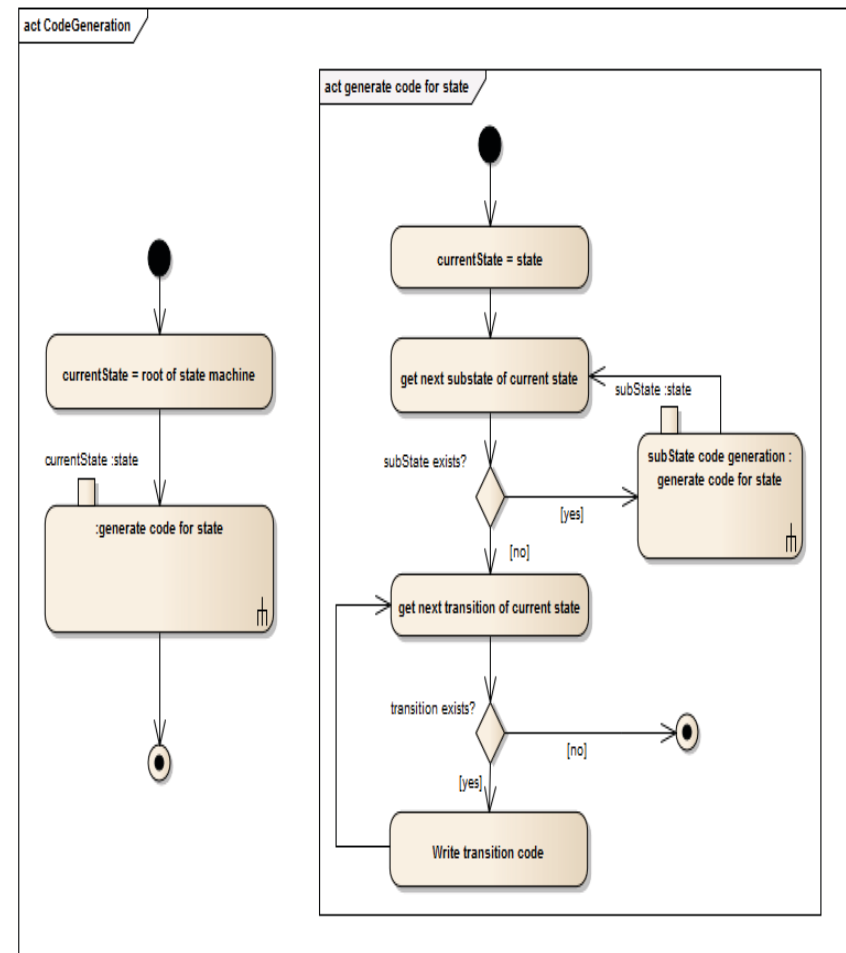
Decisions need to be made about what information the system will need to represent, and how the information held by the system corresponds to the real world phenomena being represented.

- Data modelling provides the opportunity to address these issues in RE.



RE -> Modelling and Analysing Requirements -

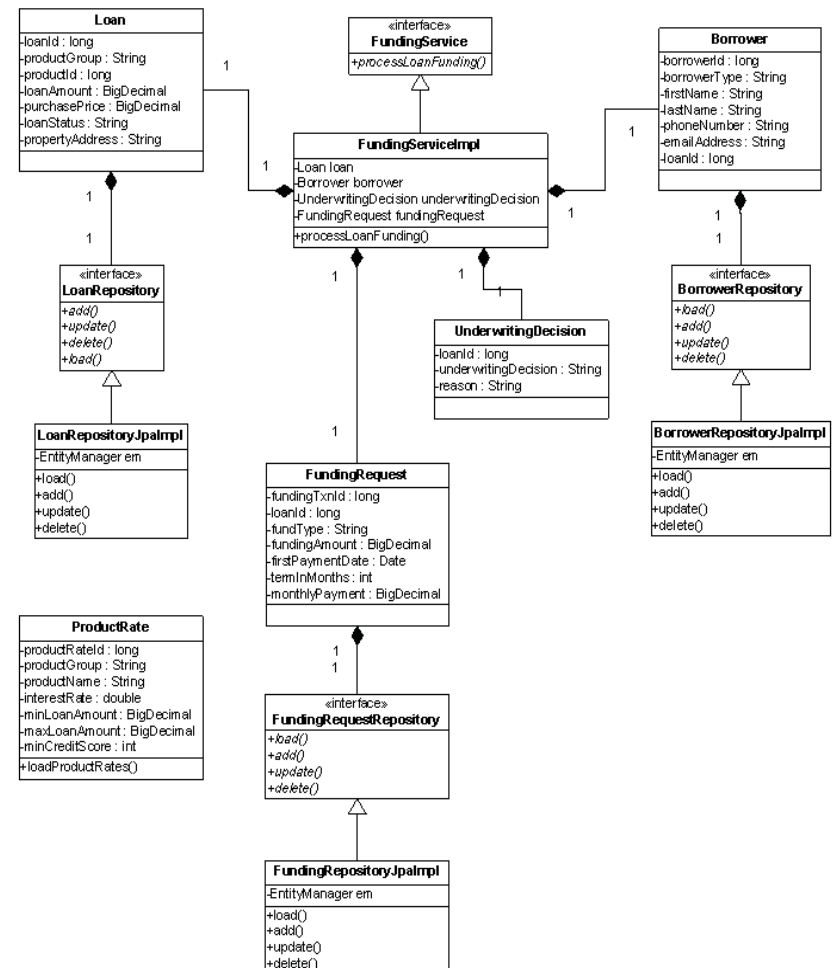
- **C) Behavioural Modelling**
 - Modelling requirements often involves modelling the dynamic or functional behaviour of stakeholders and systems, both existing and required



RE -> Modelling and Analysing Requirements -

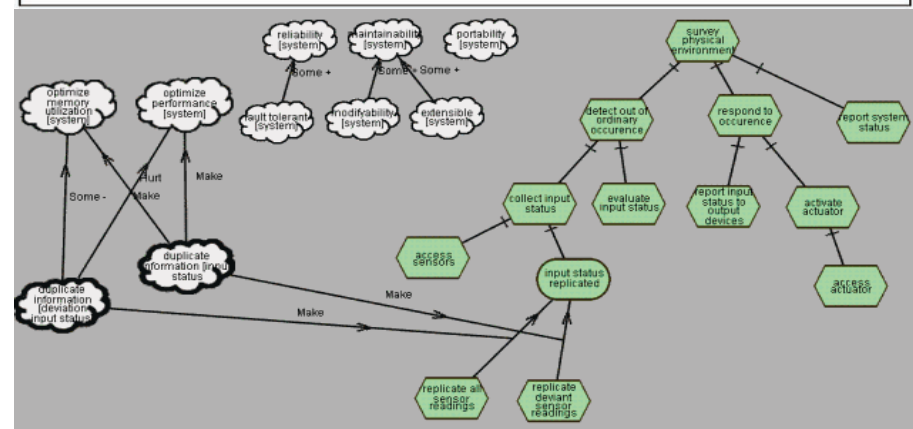
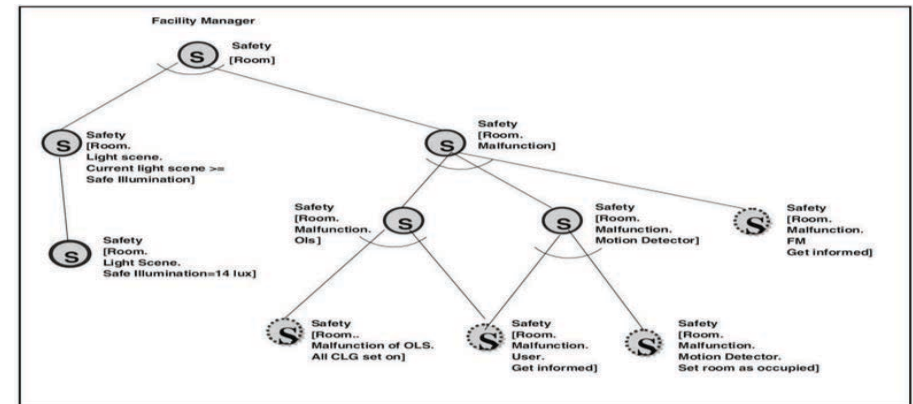
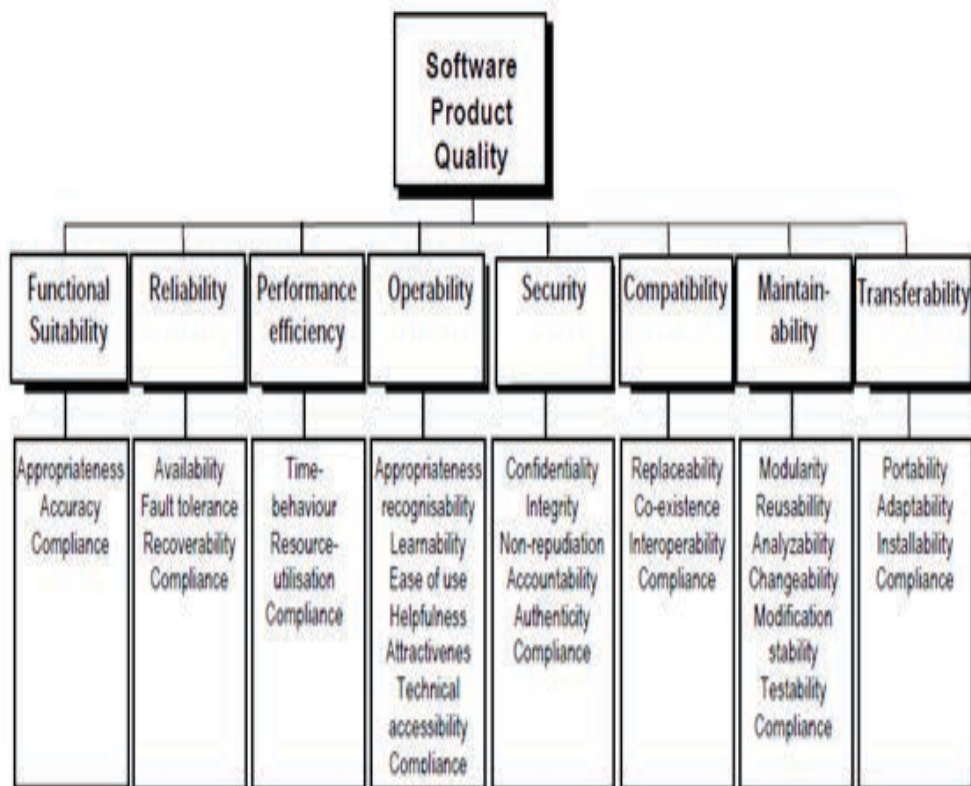
- **D) Domain Modelling** - A significant proportion of the RE process is about developing domain descriptions. A model of the domain provides an abstract description of the world in which an envisioned system will operate.

Loan Processing Application Domain Model



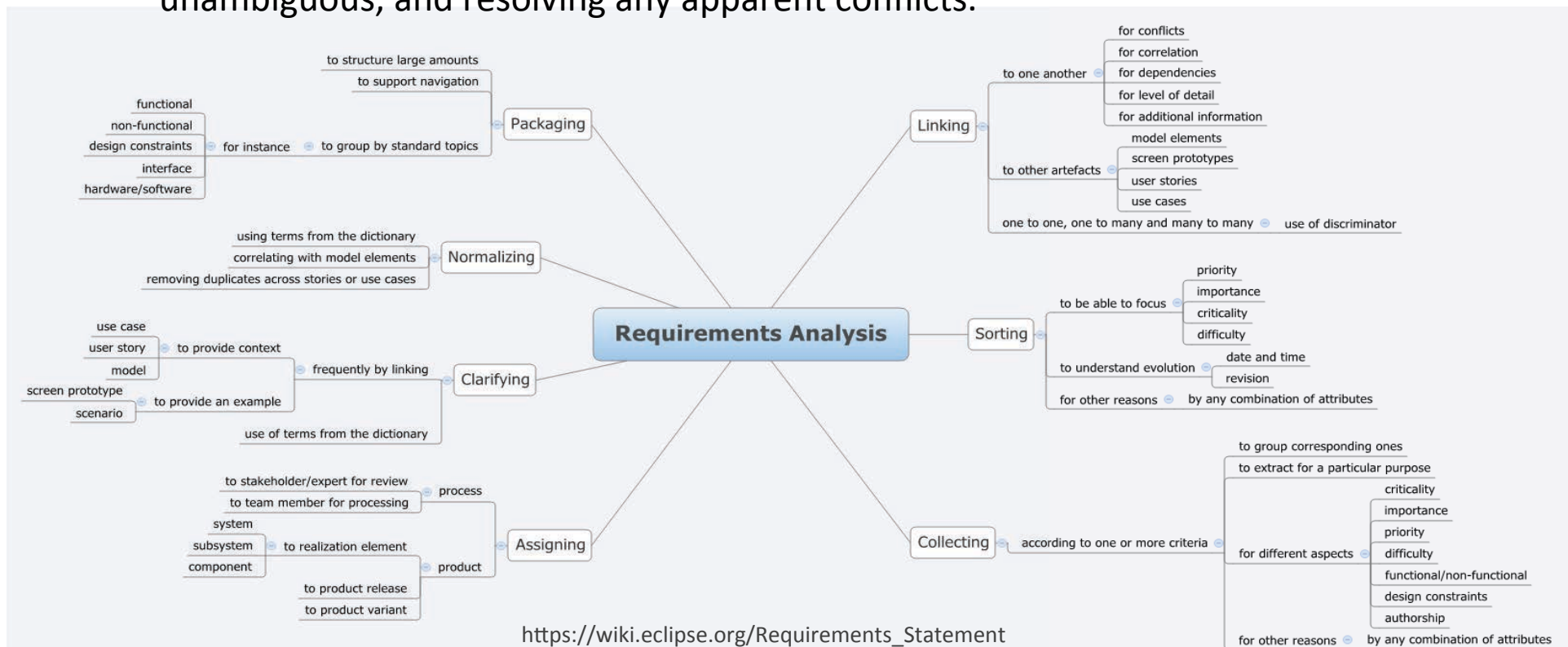
RE -> Modelling and Analysing Requirements -

- **E) Modelling Non-Functional Requirements (NFRs)**
 - Non-functional requirements (also known as quality requirements) are generally more difficult to express in a measurable way, making them more difficult to analyse

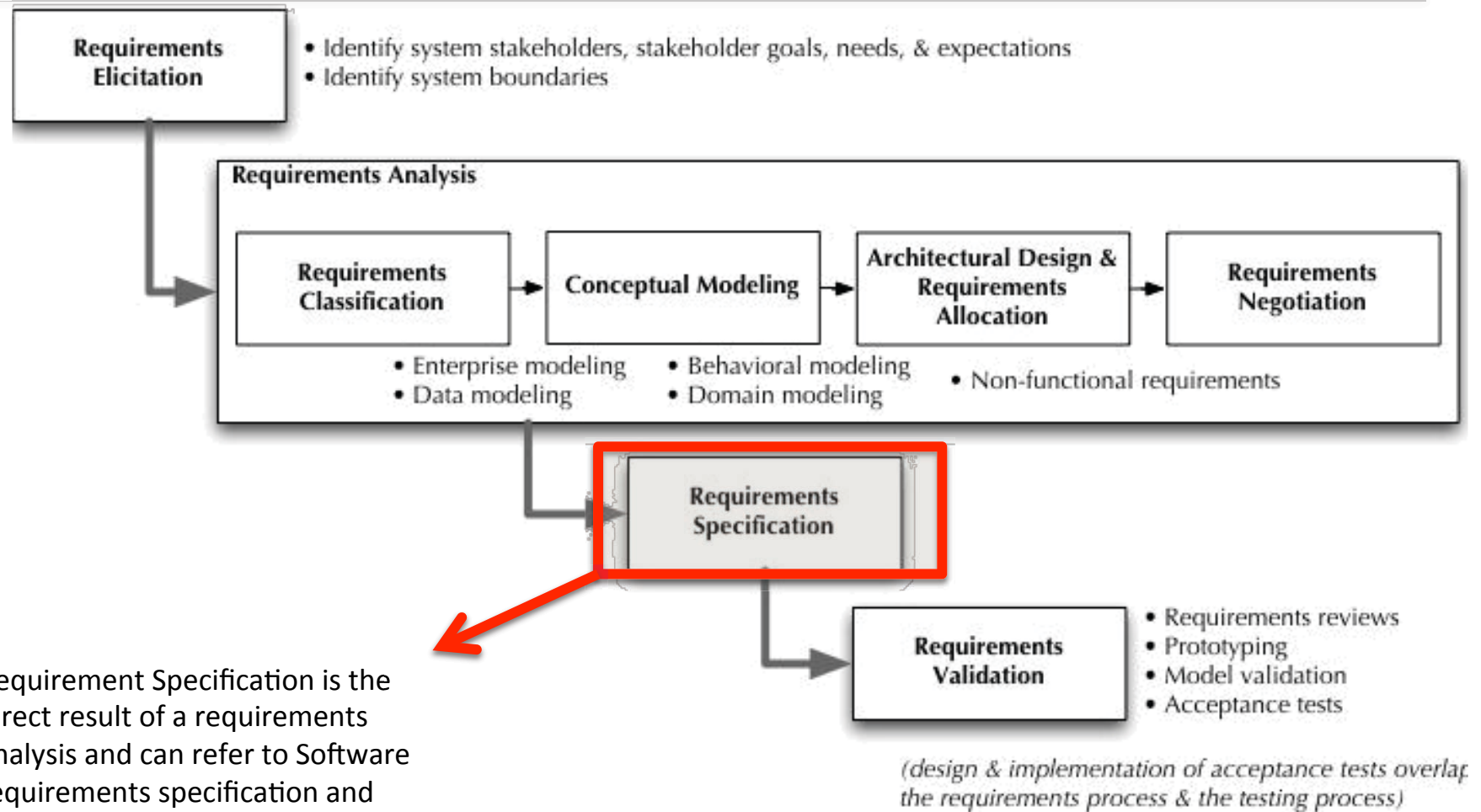


RE -> Modelling and Analysing Requirements -

- **F) Analysing Requirements Models** - A primary benefit of modelling requirements is the opportunity this provides for analysing them determining whether the stated requirements are clear, complete, consistent and unambiguous, and resolving any apparent conflicts.



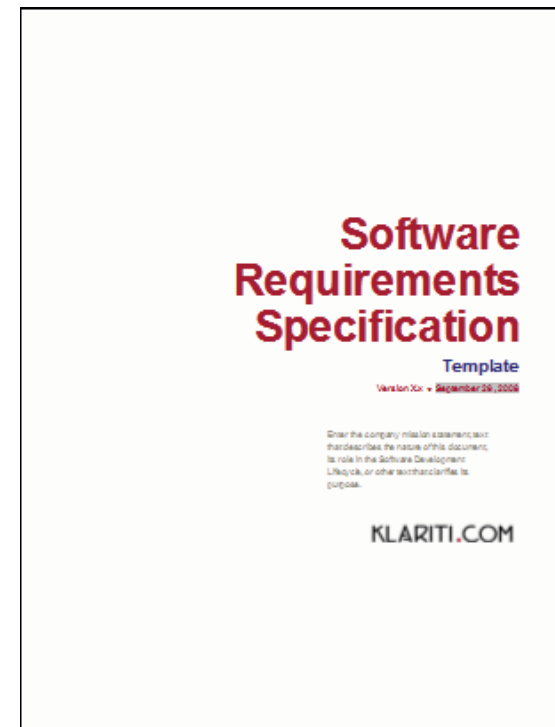
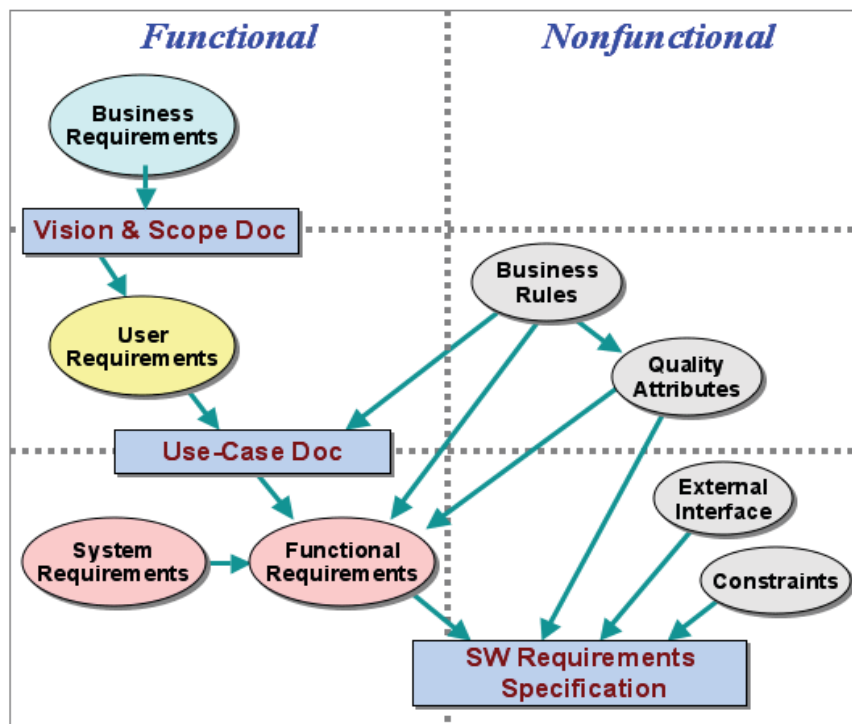
RE -> Requirement Specification



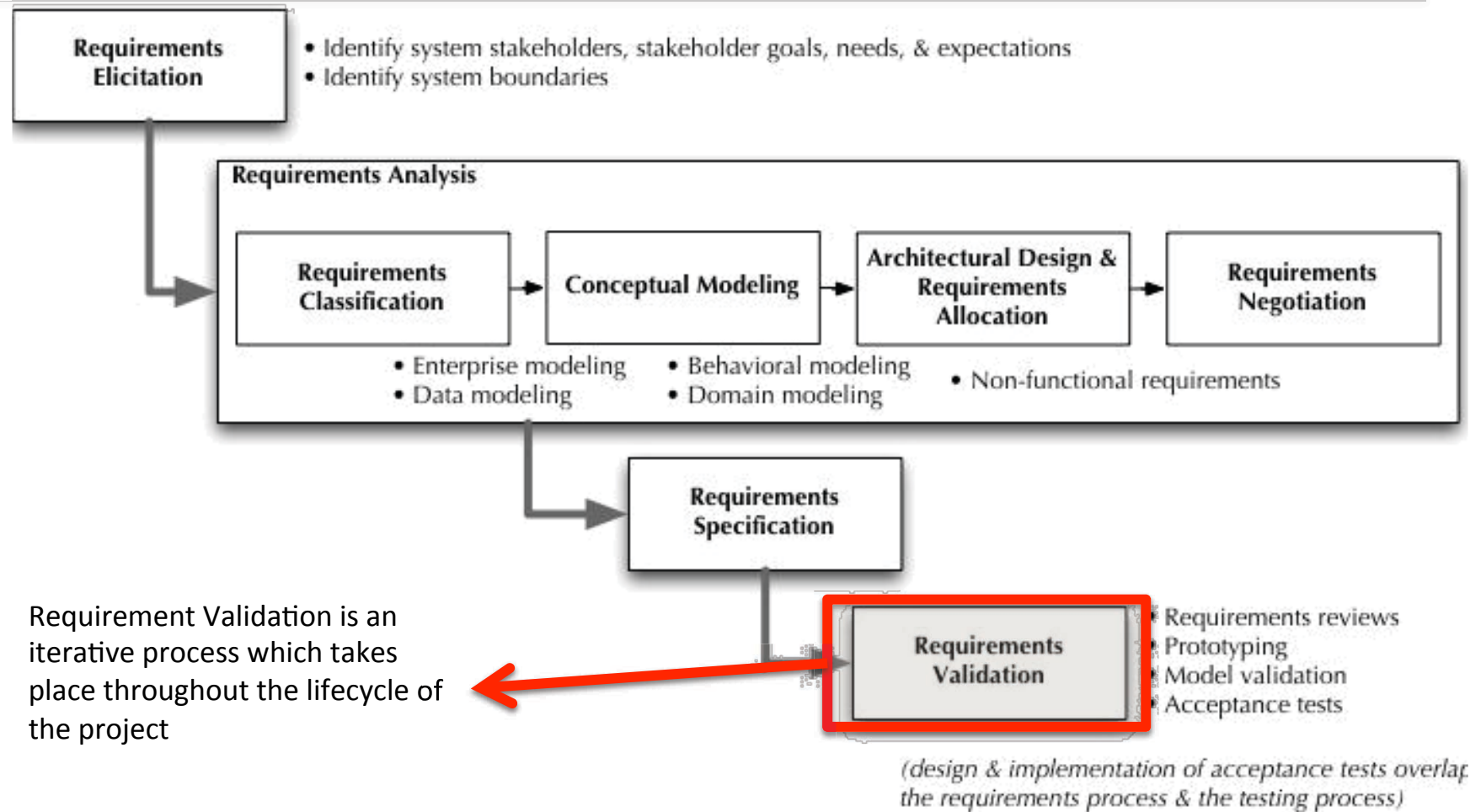
Requirement Specification is the direct result of a requirements analysis and can refer to Software requirements specification and Hardware requirements specification

RE - Software requirements specification (SRS)

- A software requirements specification (SRS) is a comprehensive description of the intended purpose and environment for software under development. The SRS fully describes what the software will do and how it will be expected to perform.
- The process of writing down the user and system requirements in a requirements document.

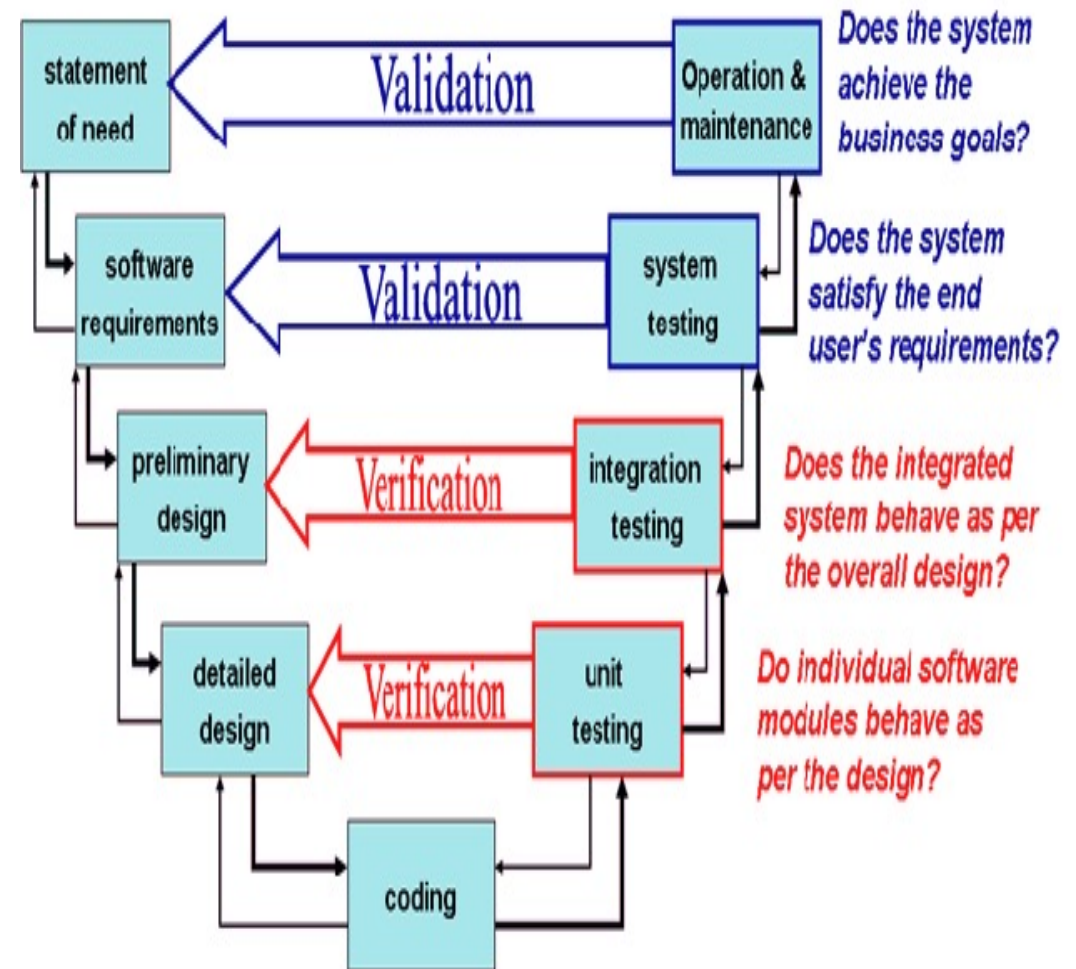


RE -> Requirements Validation



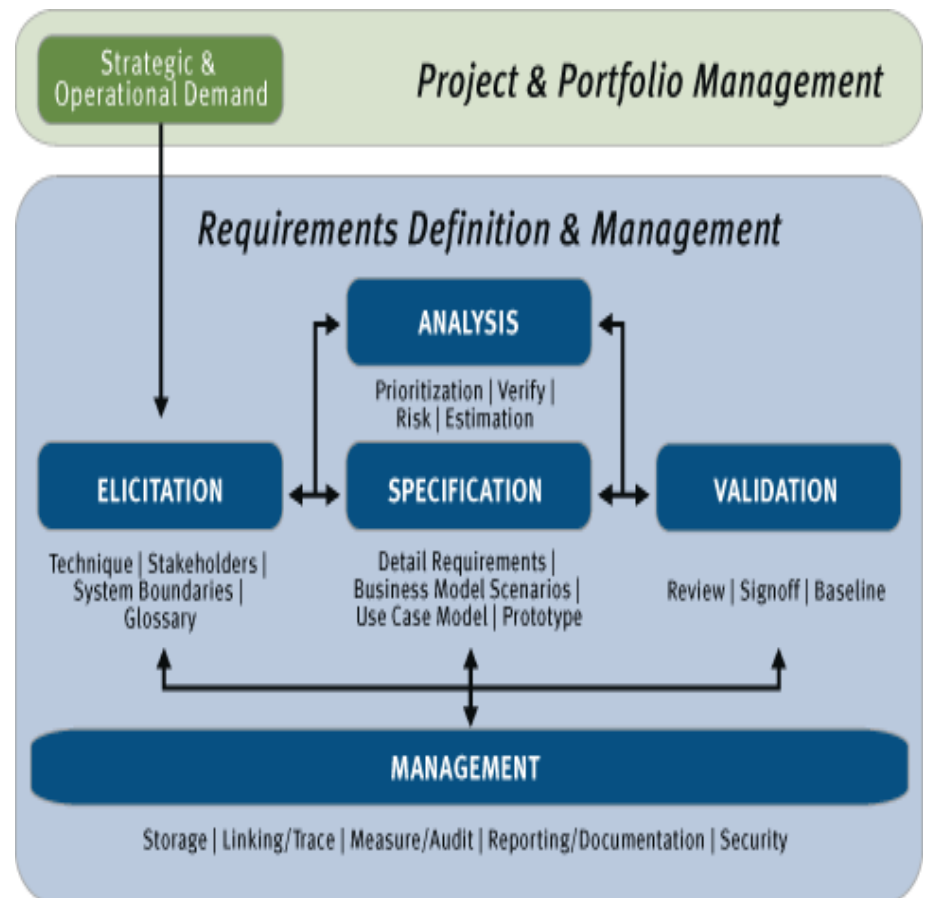
RE -> Requirements Validation

- Boehm's definition:
 - Validation: Am I building the right system?
 - Verification: Am I building the system right?
- Requirements validation is difficult for two reasons:
 - Philosophical in nature, and concerns the question of truth and what is knowable.
 - Social, and concerns the difficulty of reaching agreement among different stakeholders with conflicting goals.



RE - > Additional - Evolving Requirements

- Managing change is a fundamental activity in RE.
 - Changes to requirements documentation
 - Traceability links to monitor and control the impact of changes in different parts of the documentation.
 - Managing inconsistency in requirements specifications as they evolve is a major challenge.



RE - requirements validation session



RE Summary

1. **SOCIAL SCIENCES:** RE draws on the cognitive and social sciences:
 - Cognitive Psychology, Anthropology, Sociology, Linguistics
2. **BOUNDARIES:** The identification of stakeholders and user classes, of goals and tasks, and of scenarios and use cases all depend on how the boundaries are chosen.
3. **MODELS:** Models can be used to represent a whole range of products of the RE process.
4. **COMMUNICATION:** RE is not only a process of discovering and specifying requirements, it is also a process of facilitating effective communication of these requirements among different stakeholders.
5. **DOCUMENTATION:** The way in which requirements are documented plays an important role in ensuring that they can be read, analysed, (re-)written, and validated.
6. **VALIDATION:** Requirements validation is difficult for two reasons.
 - The first reason is philosophical in nature, and concerns the question of truth and what is knowable.
 - The second reason is social, and concerns the difficulty of reaching agreement among different stakeholders with conflicting goals.
7. **CHANGES:** Managing changing requirements is not only a process of managing documentation, it is also a process of recognising change through continued requirements elicitation, re- evaluation of risk, and evaluation of systems in their operational environment.

RE Humour



"HOW TO TORTURE A BUSINESS ANALYST # 3"



Tutorial:

Procesamiento de Lenguaje Natural en Ingeniería de Requisitos: *Contribuciones Potenciales y Desafíos de Investigación*

CIBSE 2015 – Lima, Peru

Dr Ricardo Gacitúa
ricardo.gacitua@ceisufro.cl

Centro de Estudios en Ingeniería de Software
Depto. Cs. Computación e Informática
Universidad de La Frontera
Temuco, Chile

Indice

- Ingeniería de Requerimientos (RE)
- **Procesamiento de Lenguaje Natural (NLP)**
- Avances en la relación entre RE & NLP
- Contribuciones Potenciales
- **Desafíos de Investigación**

Después de este tutorial, podrás :

- Definir RE y su describir su importancia.
- **Definir NLP y sus áreas de acción.**
- Conocer el estado del arte en la relación entre NLP y RE.
- Describir las contribuciones potenciales de NLP para RE.
- **Listar los principales desafíos para la investigación en RE.**

Can computational methods aid the understanding of human language?

中文事工部

许多华人来到加拿大定居都面临着这样的情况：自己以及我们的父母的第一语言为中文，可是我们的下一代的母语则是英语。为兼顾中英文同得事的需求，西苑教会的中文事工部因此应运而生，在伦敦华人教会中，以其独特的方式，各自用母语来敬拜赞美主，聆听来自神的教导，一起携手踏上改变生命的旅程。

中文团诗崇拜从12月4日开始，每周日上午10:45~12:10。

同时教会提供各年龄段孩子由专职老师带领的丰富多彩的团活动（英文）：

幼儿看护：学龄前儿童

KidsWalk: 1~5年级少儿活动

Transit: 6~8年级学生活动

（1~8年级的学生先和父母在一起参加音乐敬拜直至他们各自的活动开始）

中文事工部目前设有生命培训中心，生命小组，关爱小组，栽培小组等各项活动和事工。

我们欢迎您和我们一起来敬拜赞美主。如果您想更多地了解各项中文事工或参与服事，请与我们联系。

中文事工部电话：519-472-3182

e-mail: chineseministry@westpark.on.ca

生命培训中心 (Life Training Center)

神学装备、圣经查经、生活应用

时间：周日 9:00am - 10:15pm

地点：西苑教会小礼堂



Tutorial

NATURAL LANGUAGE PROCESSING

Motivation - Natural Language Processing (NLP)

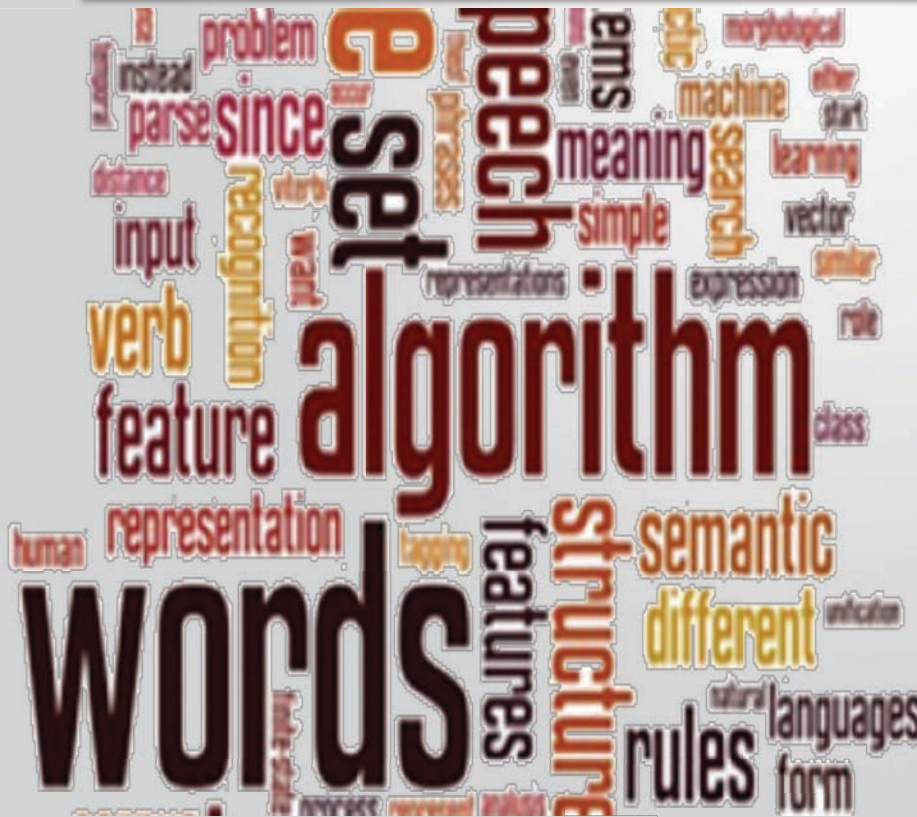
- NLP is the branch of computer science focused on developing systems that allow computers to communicate with people using everyday language.
- Also called *Computational Linguistics*
 - Also concerns how computational methods can aid the understanding of human language.



Natural Language Processing



DEFINITION Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages.

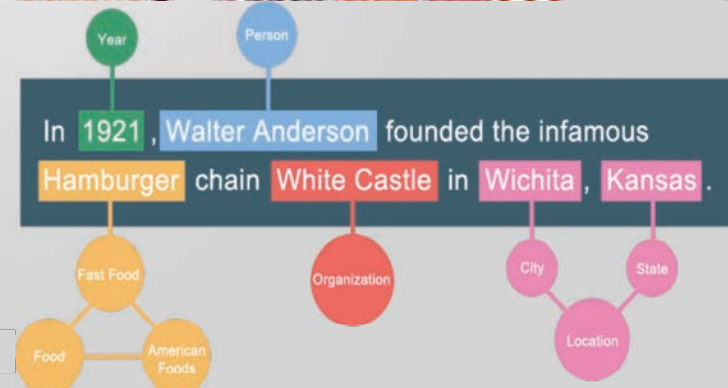
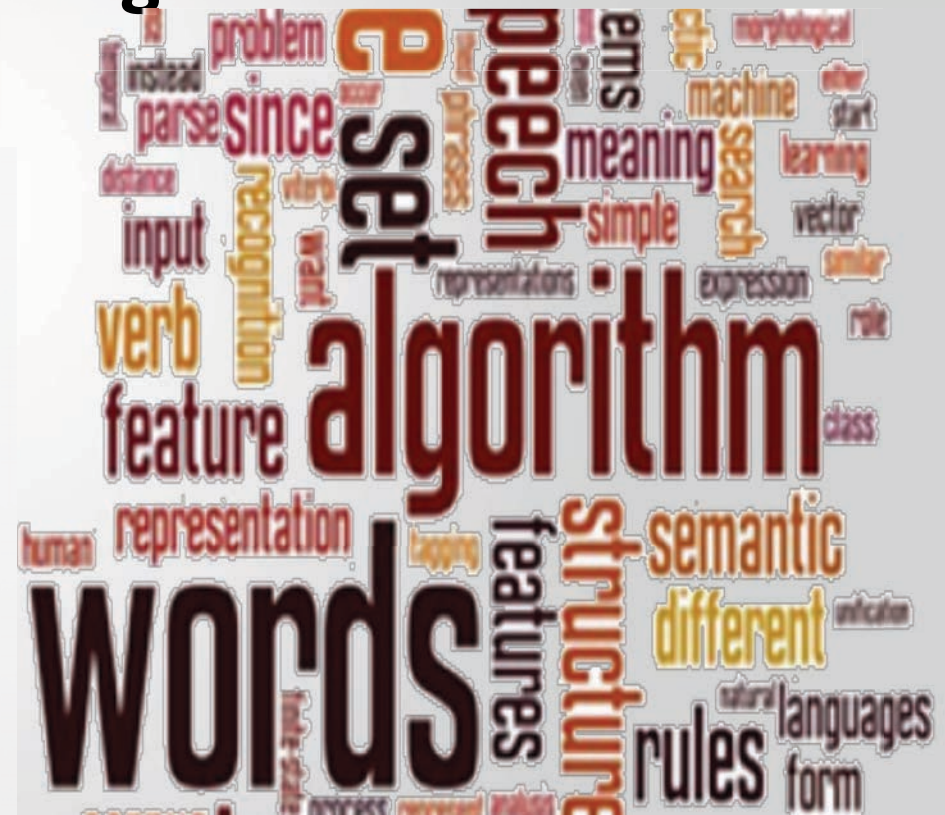


Natural Language Processing

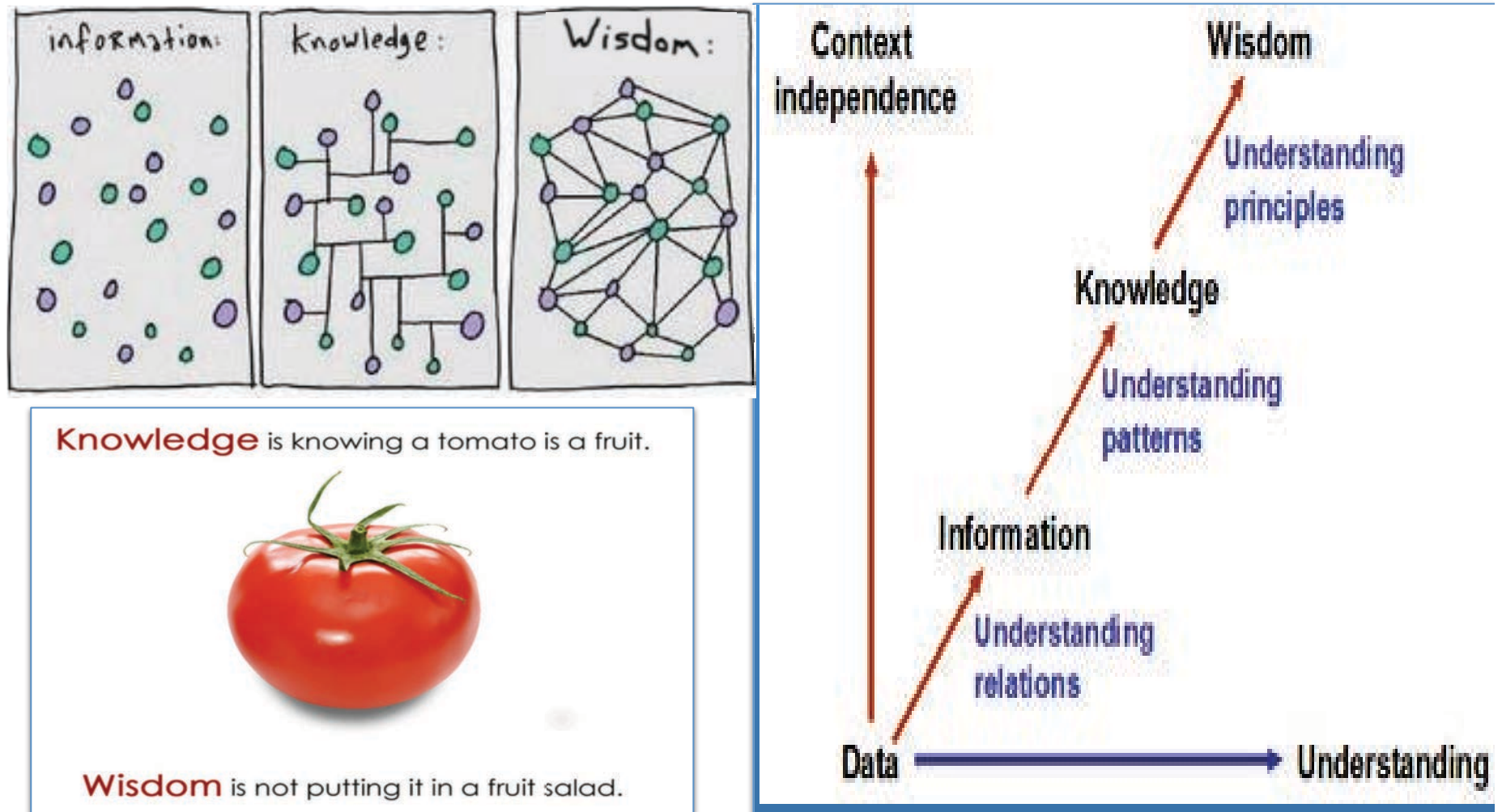
NLP is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages.

- It is the analysis of human language so that computers can understand natural languages as humans do.
- Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation.

*Christopher D. Manning and Hinrich Schütze (1999).
Foundations of Statistical Natural Language Processing.
The MIT Press.*



Information, Knowledge and Wisdom



Natural language

Conversational

Contextual

Personal

Works with built-in apps

Dictation anywhere

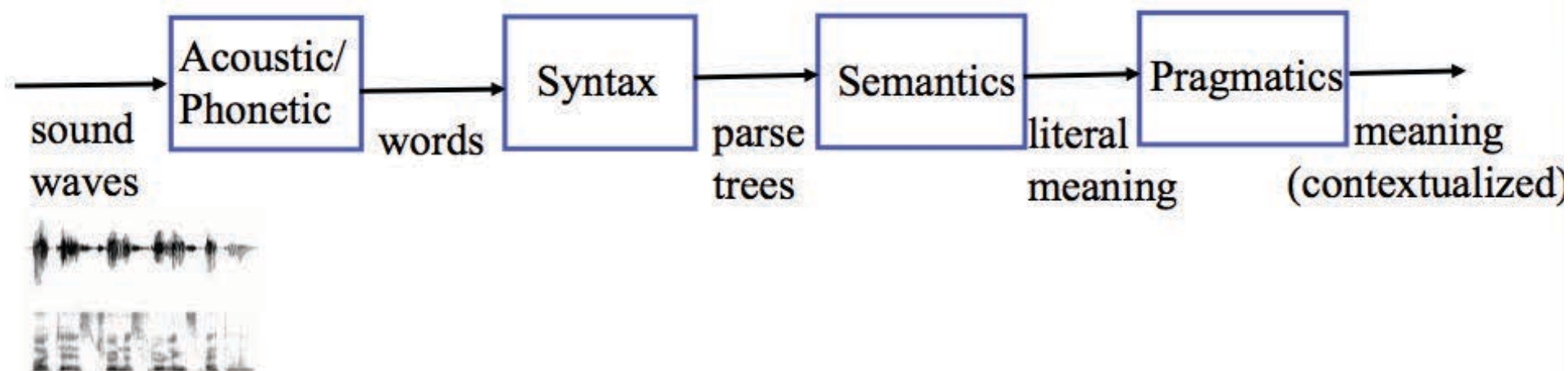
3G + Wi-Fi



Natural Language Processing

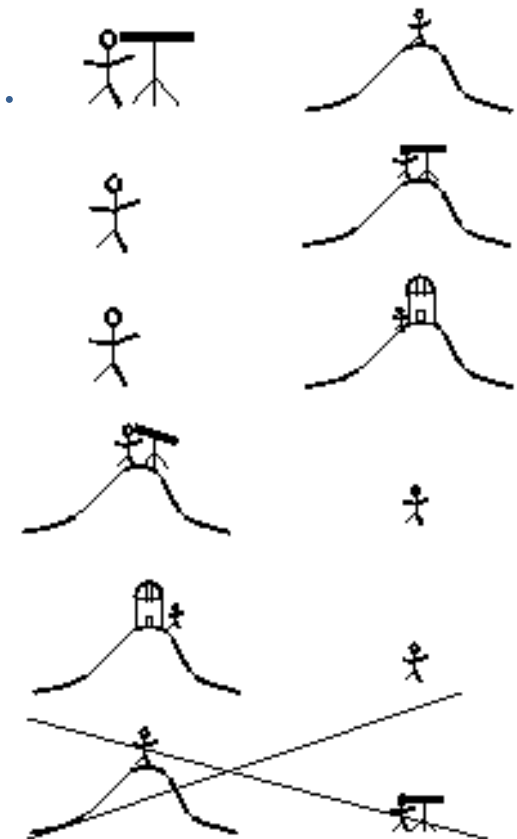


Modular Comprehension



Ambiguity

- Natural language is highly ambiguous and must be *disambiguated*.
 - I saw the man on the hill with a telescope.



Major tasks in NLP

RE	NLP		Survey	Challenges	
Automatic Summarization Produce a readable summary of a chunk of text	Co-reference Resolution Given a sentence, determine which words ("mentions") refer to the same objects ("entities")	Discourse Analysis To identifying the discourse structure of connected text,(i.e. elaboration, explanation, contrast).	Machine Translation Automatically translate text from one human language to another.	Part-of-speech tagging Given a sentence, determine the part of speech for each word	Parsing Determine the parse tree (grammatical analysis) of a given sentence
Morphological Segmentation Separate words into individual morphemes and identify the class of the morphemes.	Named Entity Recognition (NER) Given a stream of text, determine which items in the text map to proper names	Natural Language Understanding Convert chunks of text into more formal representations such as first-order logic	Optical character recognition Given an image representing printed text, determine the corresponding text.	Question Answering Given a human-language question, determine its answer	Relationship Extraction Given a chunk of text, identify the relationships among named entities
Sentence breaking Given a chunk of text, find the sentence boundaries.	Sentiment Analysis Extract subjective information to determine "polarity" about specific objects	Word sense disambiguation Many words have more than one meaning; we have to select the meaning which makes the most sense in context	Speech Segmentation Given a sound clip of a person or people speaking, separate it into words	Topic Segmentation and recognition Given a chunk of text, separate it into segments each of which is devoted to a topic.	Word Segmentation Separate a chunk of continuous text into separate words.
Speech Recognition Given a sound clip of a person speaking, determine the textual representation of the speech	Information Retrieval This is concerned with storing, searching and retrieving information		Information Extraction This is concerned in general with the extraction of semantic information from text		

RE

NLP

Survey

Challenges

Automatic Summarization

Produce a readable summary of a chunk of text

Coreference Resolution

Given a sentence, determine which words ("mentions") refer to the same objects ("entities")

Discourse Analysis

To identifying the discourse structure of connected text, i.e. elaboration, explanation, contrast)."

Machine Translation

Automatically translate text from one human language to another.

Part-of-speech tagging

Given a sentence, determine the part of speech for each word

Parsing

Determine the parse tree (grammatical analysis) of a given sentence

Morphological Segmentation

Separate words into individual morphemes and identify the class of the morphemes.

Named Entity Recognition (NER)

Given a stream of text, determine which items in the text map to proper names

Natural Language Understanding

Convert chunks of text into more formal representations such as first-order logic

Optical character recognition

Given an image representing printed text, determine the corresponding text.

Question Answering

Given a human-language question, determine its answer

Relationship Extraction

Given a chunk of text, identify the relationships among named entities

Sentence breaking

Given a chunk of text, find the sentence boundaries.

Sentiment Analysis

Extract subjective information to determine "polarity" about specific objects

Word sense disambiguation

Many words have more than one meaning; we have to select the meaning which makes the most sense in context

Speech Segmentation

Given a sound clip of a person or people speaking, separate it into words

Topic Segmentation and recognition

Given a chunk of text, separate it into segments each of which is devoted to a topic.

Word Segmentation

Separate a chunk of continuous text into separate words.

Speech Recognition

Given a sound clip of a person speaking, determine the textual representation of the speech

Information Retrieval

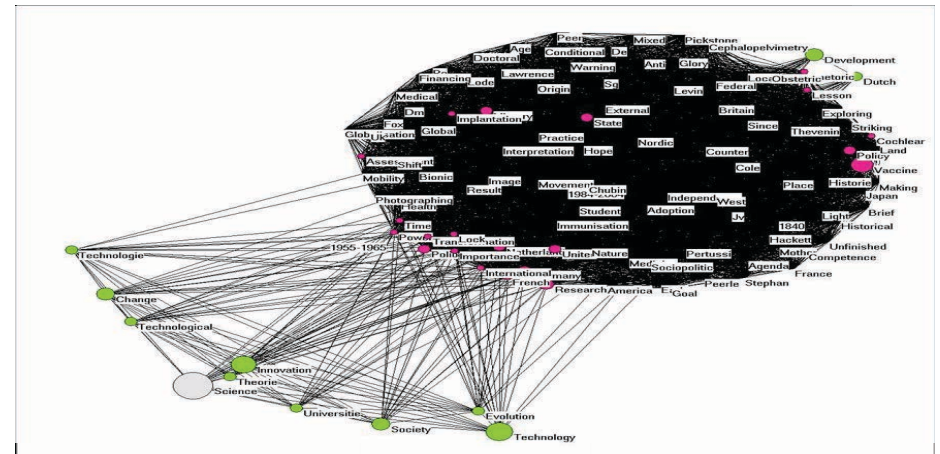
This is concerned with storing, searching and retrieving information

Information Extraction

This is concerned in general with the extraction of semantic information from text

NLP-> Automatic Discourse Analysis

- “The study of well-established meanings or ideas around a topic which shape how we can [think and] talk about it.” Stephanie Taylor’s (2010)
- For example: Segmentation by thematic



BACKGROUND KNOWLEDGE:

Recent studies indicate ...
... the previously proposed ...
... is universally accepted ...

CONTRASTING IDEAS:

... unorthodox view resolves ...
paradoxes ...
In contrast with previous hypotheses ...
... inconsistent with past findings ...

GENERALIZING:

... emerging as a promising approach
Our understanding ... has grown
exponentially ...
... growing recognition of the
importance ...

NOVELTY:

... new insights provide direct
evidence we suggest a new ...
approach ...
... results define a novel role ...

SIGNIFICANCE:

studies ... have provided important
advances
Knowledge ... is crucial for ...
understanding
valuable information ... from studies

SURPRISE:

We have recently observed ...
surprisingly
We have identified ... unusual
The recent discovery ... suggests
intriguing roles

OPEN QUESTION:

... little is known ...
... role ... has been elusive
Current data is insufficient ...

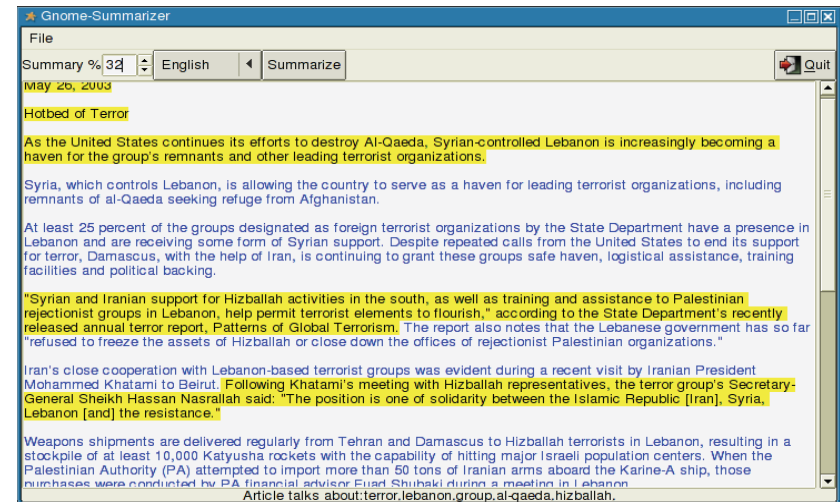
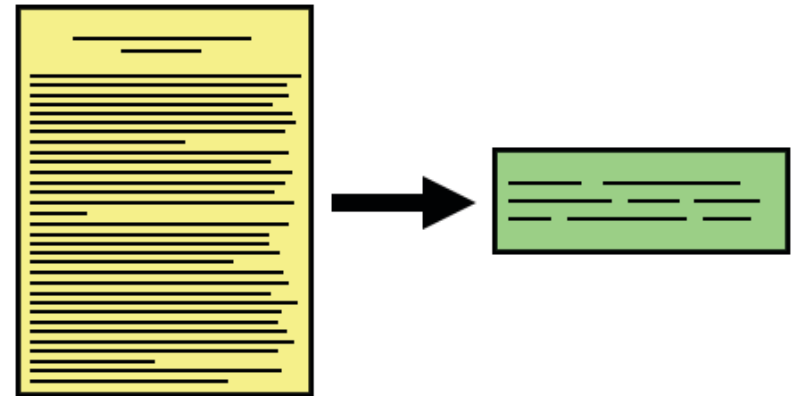
SUMMARIZING:

The goal of this study ...
Here, we show ...
Altogether, our results ... indicate

NLP -> Text Summarization

- Text summarisation
 - a **reductive transformation** of source text to summary text through content reduction by **selection** and/or **generalisation** on what is **important** in the source. (Sparck Jones, 1999)
 - the process of **distilling** the most important information from a source (or sources) to produce an abridged version for a **particular user (or users) and task (or tasks)**. (Mani and Maybury, 1999)

Automatic text summarisation = The process of producing summaries automatically.



NLP - > Machine Translate

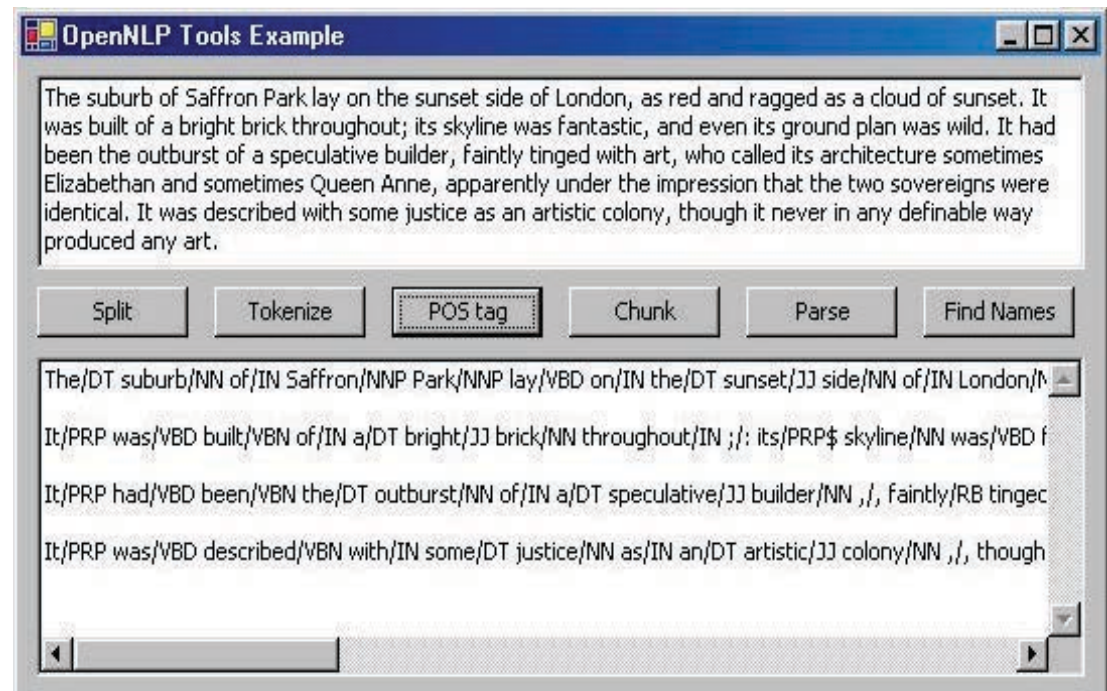
- Automatically translate text from one human language to another.
 - On a basic level, MT performs simple substitution of words in one natural language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed.
 - Solving this problem with corpus and statistical techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies.



NLP - > Part-of-Speech Tagging

POS Tagging

- It is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context—i.e. relationship with adjacent and related words in a phrase, sentence, or paragraph.
 - A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc.



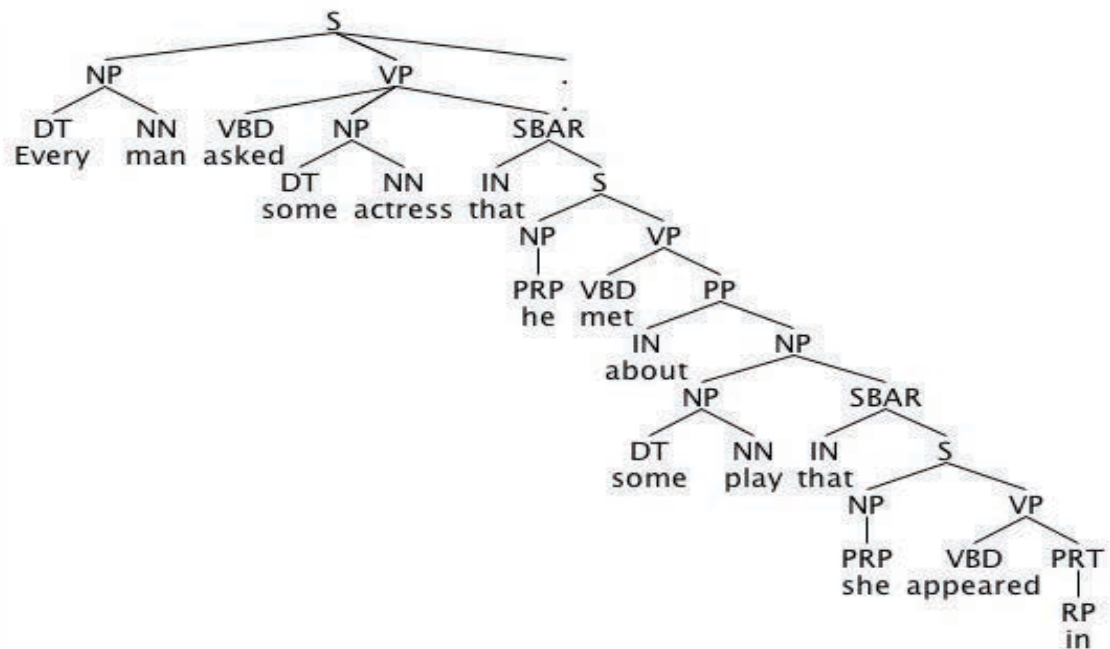
NLP - > Parsing

Parsing

Parsing or syntactic analysis is the process of analysing a string of symbols, either in natural language or in computer languages, according to the rules of a formal grammar.

Sentence Count: 317 Displayed Tree (Sentence): 287

((I (PRP)) (am (VBP)) (proud (JJ)) (that (IN)) (John (NNP))
 ((John (NNP)) (is (VBZ)) (tough (JJ)) (to (TO)) (please (VI
 ((It (PRP)) (is (VBZ)) (tough (JJ)) (to (TO)) (please (VB)) (
 ((John (NNP)) (is (VBZ)) (likely (JJ)) (to (TO)) (win (VB)) (
 ((Who (PRP)) (does (VBZ)) (his (PRP\$)) (mother (NN)) (I
 ((His (PRP\$)) (mother (NN)) (loves (VBZ)) (everyone (NI
 ((Every (RB)) (man (NN)) (likes (NNS)) (some (DT)) (syn
 ((Every (RB)) (man (NN)) (asked (VBD)) (some (DT)) (ai
 ((Someone (NN)) (gave (VBD)) (every (JJ)) (actress (NN)
 ((Which (NNP)) (man (NN)) (liked (VBD)) (which (WDT))
 ((John (NNP)) (seems (VBZ)) (that (IN)) (he (PRP)) (likes
 ((John (NNP)) (asked (VBD)) (Mary (NNP)) (about (IN)) i
 ((We (PRP)) (like (IN)) (myself (PRP)))
 ((John (NNP)) (seems (VBZ)) (is (VBZ)) (crazy (NN)))
 ((John (NNP)) (tried (VBD)) (Bill (NNP)) (to (TO)) (seem
 ((John (NNP)) (tried (VBD)) (to (TO)) (be (VB)) (arrested
 ((Mary (NNP)) (is (VBZ)) (proud (VBN)) (of (IN)) (Bill (NN
 ((John (NNP)) (seems (VBZ)) (that (IN)) (he (PRP)) (is (V
 ((The (DT)) (destruction (NN)) (Rome (NNP)) (worked (V
 ((The (DT)) (destruction (NN)) (of (IN)) (Rome (NNP)) (i
 ((The (DT)) (belief (NN)) (John (NNP)) (to (TO)) (be (VB
 ((The (DT)) (belief (NN)) (of (IN)) (John (NNP)) (to (TO))
 ((The (DT)) (belief (NN)) (that (IN)) (John (NNP)) (is (VB
 ((John (NNP)) ('s (POS)) (belief (NN)) (to (TO)) (be (VB))
 ((John (NNP)) (seems (VBZ)) (that (IN)) (his (PRP\$)) (be



NLP -> Named Entity Recognition (NER)

Automatically find names
of people, places, products,
and organizations in text



In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced using the new discovery of nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell-Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.

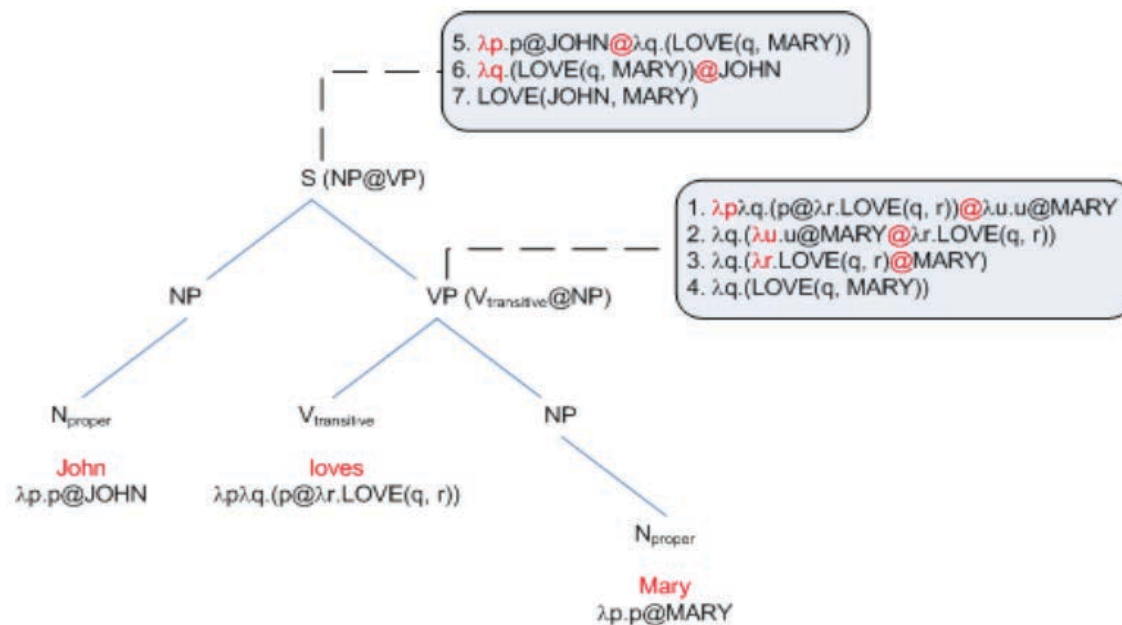
Tag colours:

LOCATION TIME PERSON ORGANIZATION MONEY PERCENT DATE

NLP -> Natural Language Understanding

- Convert chunks of text into more formal representations such as first-order logic.

(1) *John loves Mary*



NLP -> Natural Language understanding

After two years of online dating,
John finally found the courage
to formalize his relationship...


$$\exists x \mid \forall x \in \Omega, x = \text{myself},$$
$$\exists y \mid \forall y \in$$

B. Karpman 2006

NLP -> Question Answering

- **Question Answering (QA)** is a computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language.

'William Wilkinson's 'An account of the principalities of Wallachia and Moldavia' inspired this author's most famous novel

Bram Stoker



Computer Wins on 'Jeopardy!': Trivial, It's Not



Watson showed itself to be imperfect, but researchers at I.B.M. and other companies are already developing uses for Watson's technologies that could have a significant impact on the way doctors practice and consumers buy products.

NLP -> Question Answering

June 8, 2011 10:01 PM
Julie Klein



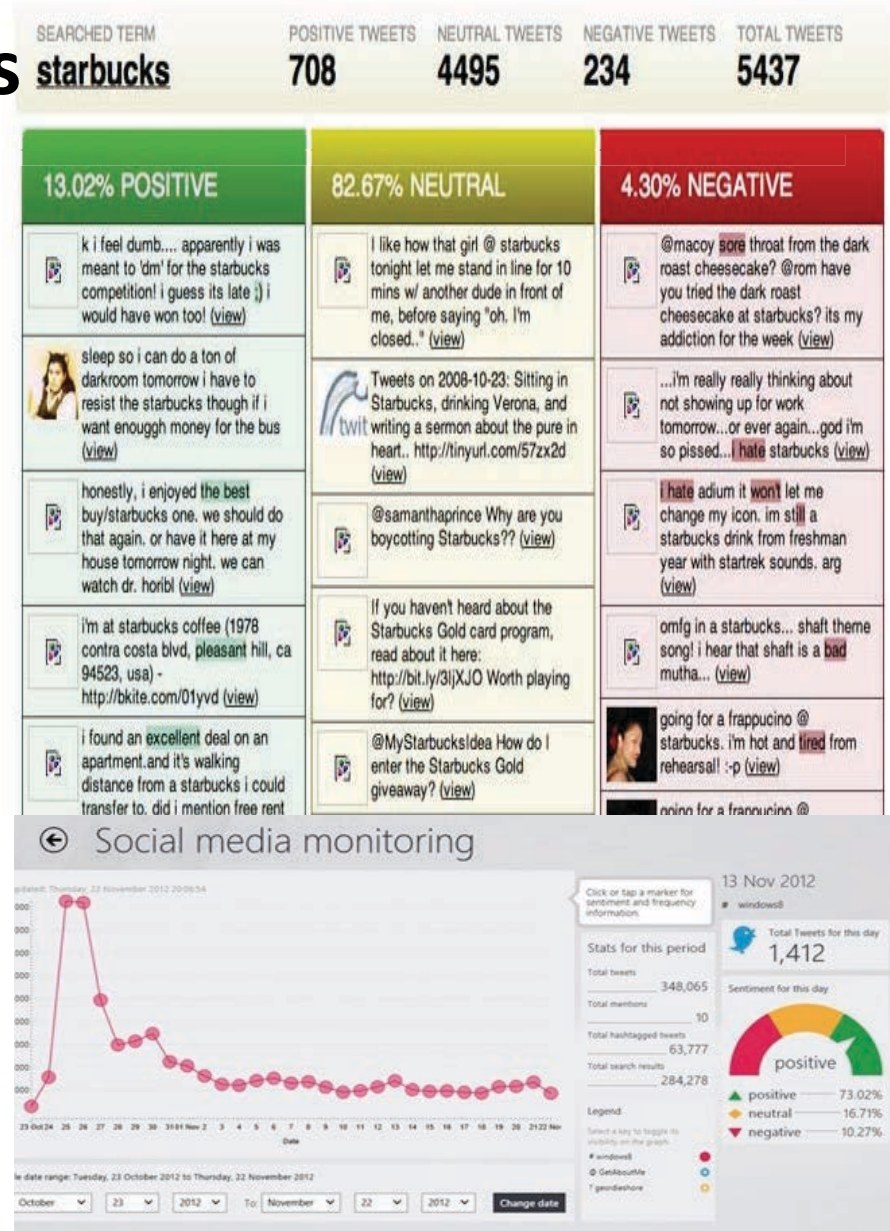
VirtuOz, creator of intelligent virtual agents that provide companies with customer service and tech support, has raised \$7M in a new round of funding. The Emeryville, Calif. company, which was founded in Paris, plans to use the investment for growth in both the United States and Europe.

The company's virtual agents, or chatbots, are programmed to automate sales support for large and mid-sized businesses. By helping online shoppers resolve issues or make purchase decisions, VirtuOz helps web sites cut labor costs. The company claims that its agents, equipped with natural language processing (NLP) capabilities, can provide customer service to clients for one-tenth the cost of a traditional human support team.



NLP – Sentiment Analysis

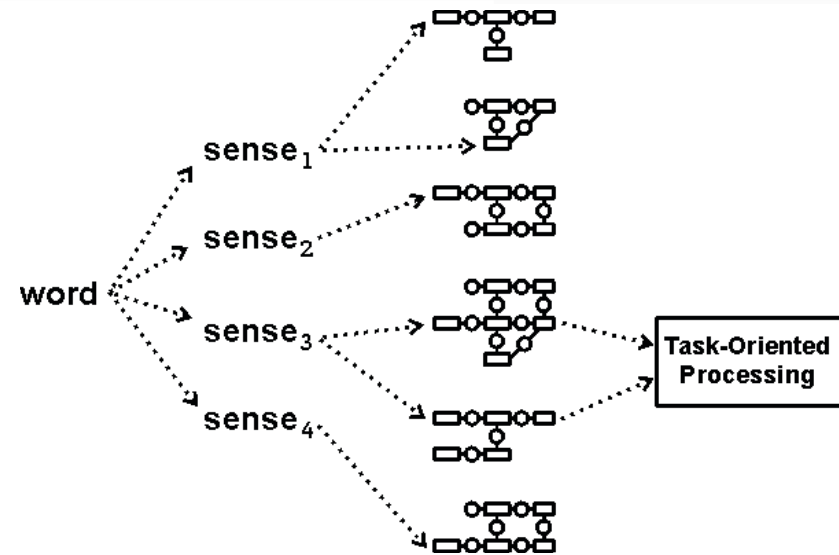
- It aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document.
- The attitude may be his or her judgment or evaluation (see appraisal theory), affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader)



Word Sense Disambiguation

WSD

- It is the process of identifying which sense of a word (i.e. meaning) is used in a sentence, when the word has multiple meanings.
- Word-sense disambiguation (WSD) is an open problem of natural language processing and ontology.



Bass: fish



???



Bass: instrument

information extraction

One stop solution to extract structured information from unstructured data.



Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents.

Subject : curriculum meeting
Date : January 21, 2014

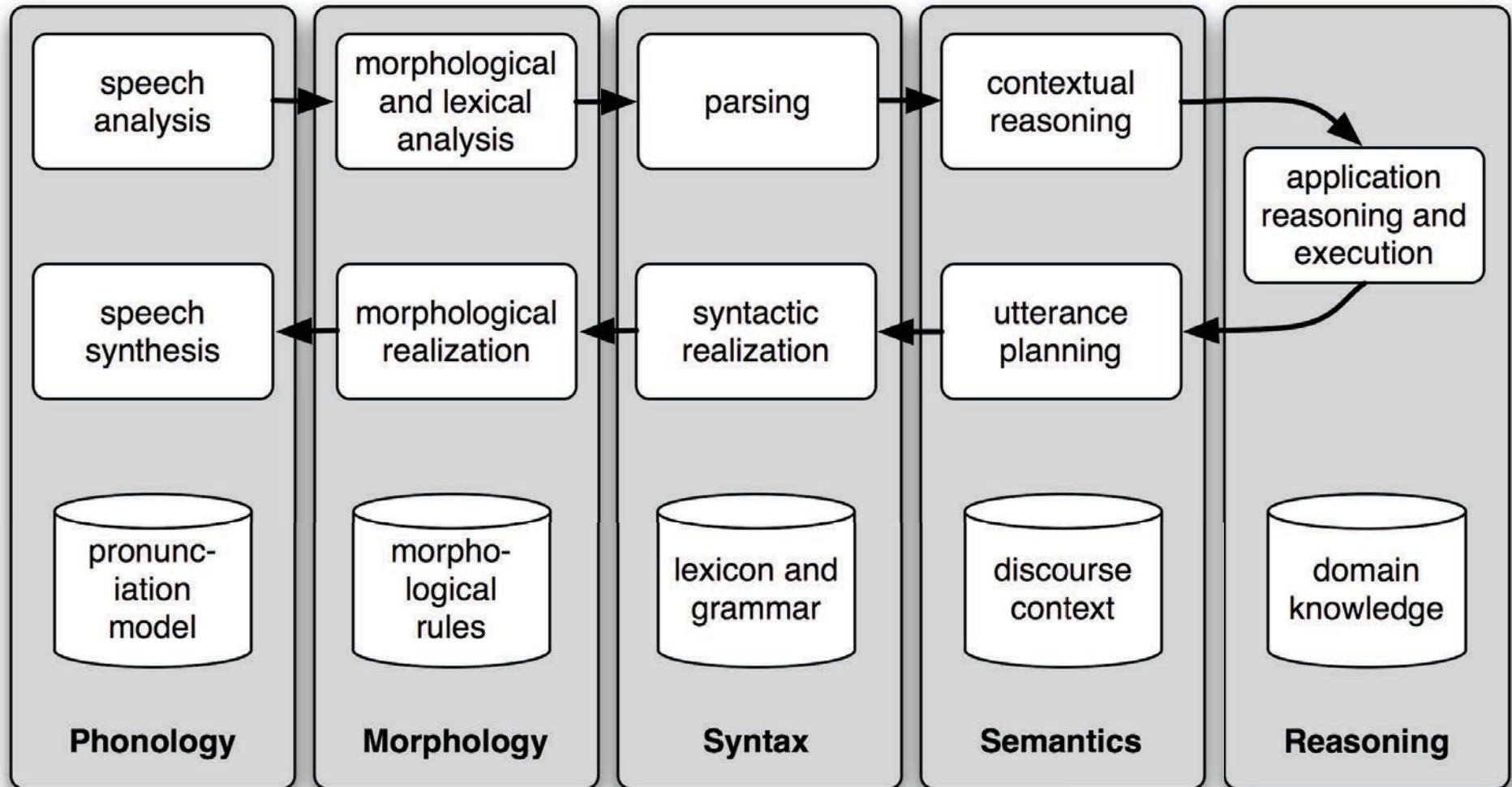
To: Ricardo Gacitua

Event: Curriculum mtg.
Date : Jan-22-2014
Start : 10:00 am
End : 11:30
Where: gates 171

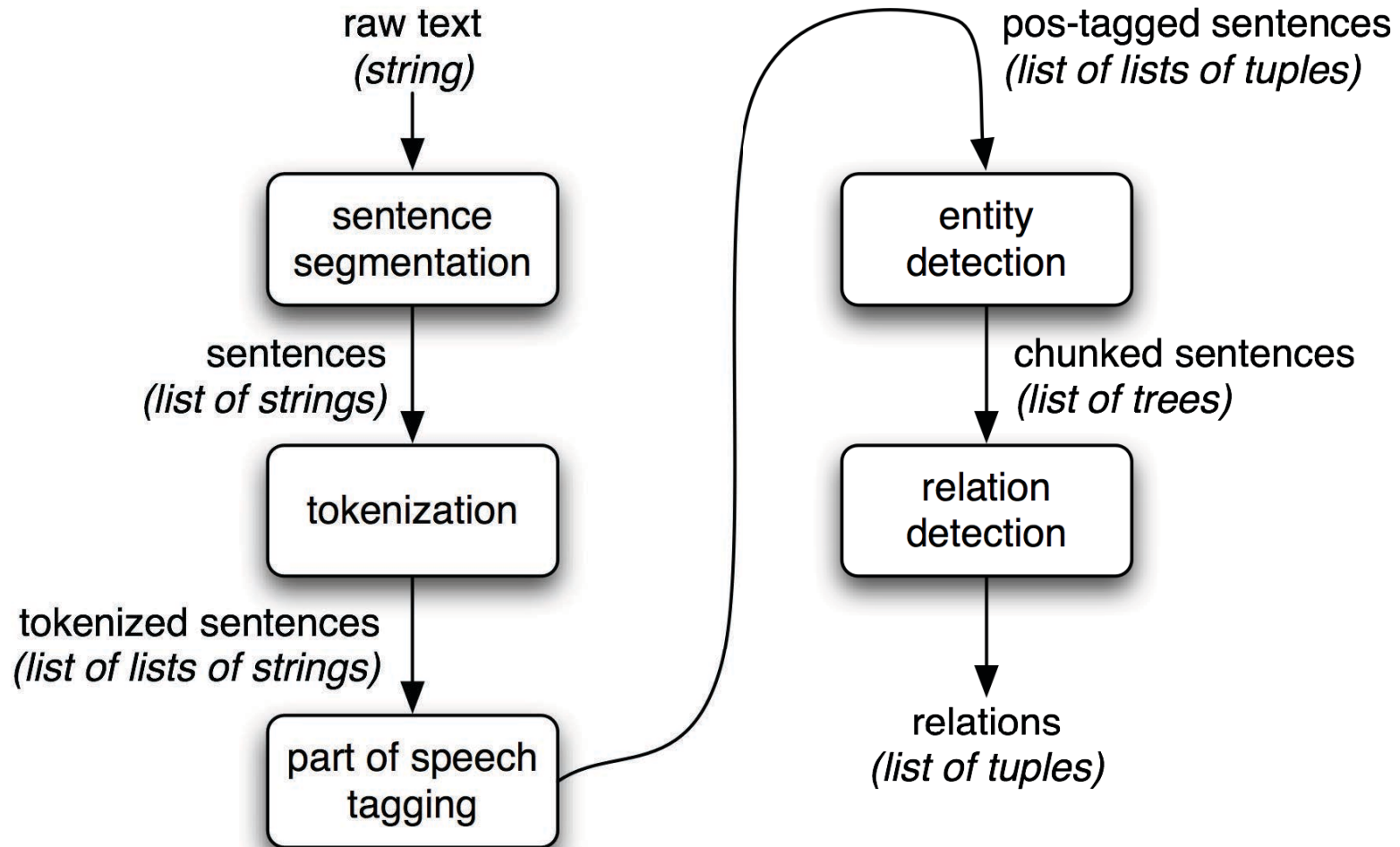
Hi Ricardo, we've now scheduled the curriculum meeting.
It will be in Gates 171 tomorrow from 10:00 – 11:30.
Carlos.

Create new Calendar Entry

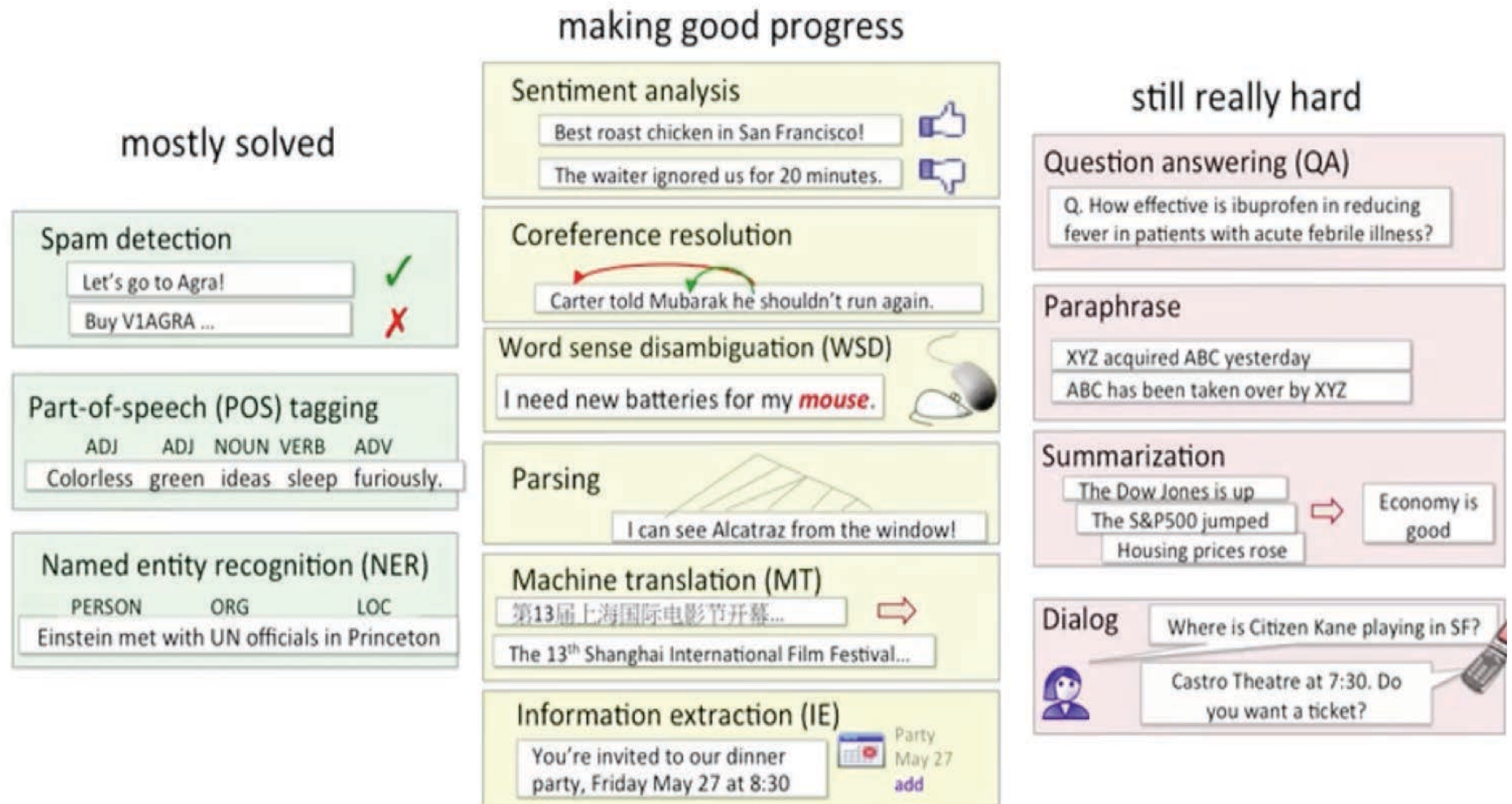
Examples - Pipeline



Example – Relation Detection



State of the art in language technologies



Tagging

- A tag is a non-hierarchical keyword or term assigned to a piece of information (such as an Internet bookmark, digital image, or computer file).
- This kind of metadata helps describe an item and allows it to be found again by browsing or searching.

directed and color Doppler studies of gallbladder tumors.

in cases of primary adenocarcinoma of the gallbladder (GB), 1 of malignant fibrous histiocytoma, 3 of metastatic adenocarcinoma, 5 of adenoma, 5 of polypus, 2 of xanthogranuloma, 6 of chronic cholecystitis, 4 of acute cholecystitis, and 8 of subacute cholecystitis were studied by image-directed and color Doppler ultrasonography (CDUS).

he 14 cases of primary GB cancer (10 masses, 4 thickening wall) were found to have a high velocity arterial blood flow signal in the wall of the GB. In contrast, the 3 cases of metastatic cancer of the GB had no blood flow signal in the wall of the GB.

30 cases of benign lesions of the GB, only in 12 cases was a low velocity blood flow signal found.

of 10 cases of primary GB malignancy were found to have high velocity arterial blood flow signals in the tumor masses. A blood flow signal was observed in the masses of 13 cases (3 of metastatic adenocarcinoma, 5 of adenoma, 5 of polypus).

normal high velocity arterial blood flow signal observed within masses in the GB or in the GB wall is a significant feature of primary GB cancer and thus can differentiate primary GB cancer from metastatic and benign lesions of the GB.

© Ricardo Gacitúa

distributed under Creative Commons
Attribution License

©Universidad de La Frontera

Tagging



"Now! ... *That* should clear up a few things around here!"

Statistical NLP

- Statistical natural-language processing uses stochastic, probabilistic and statistical methods to resolve some of the difficulties discussed above, especially those which arise because longer sentences are highly ambiguous when processed with realistic grammars, yielding thousands or millions of possible analyses

NLP Summary

- Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation.
- NLP research is gradually shifting from lexical semantics to compositional semantics and, further on, narrative understanding.
 - Human-level natural language processing, however, is an AI-complete problem. That is, it is equivalent to solving the central artificial intelligence problem—making computers as intelligent as people, or strong AI.
 - NLP's future is therefore tied closely to the development of AI in general.

Tutorial:

Procesamiento de Lenguaje Natural en Ingeniería de Requisitos: *Contribuciones Potenciales y Desafíos de Investigación*

CIBSE 2015 – Lima, Peru

Ricardo Gacitúa

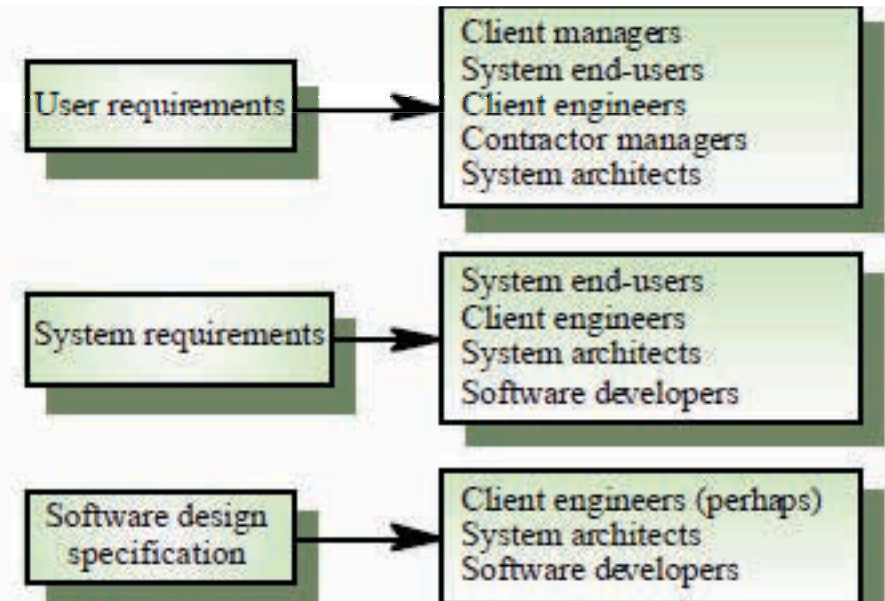
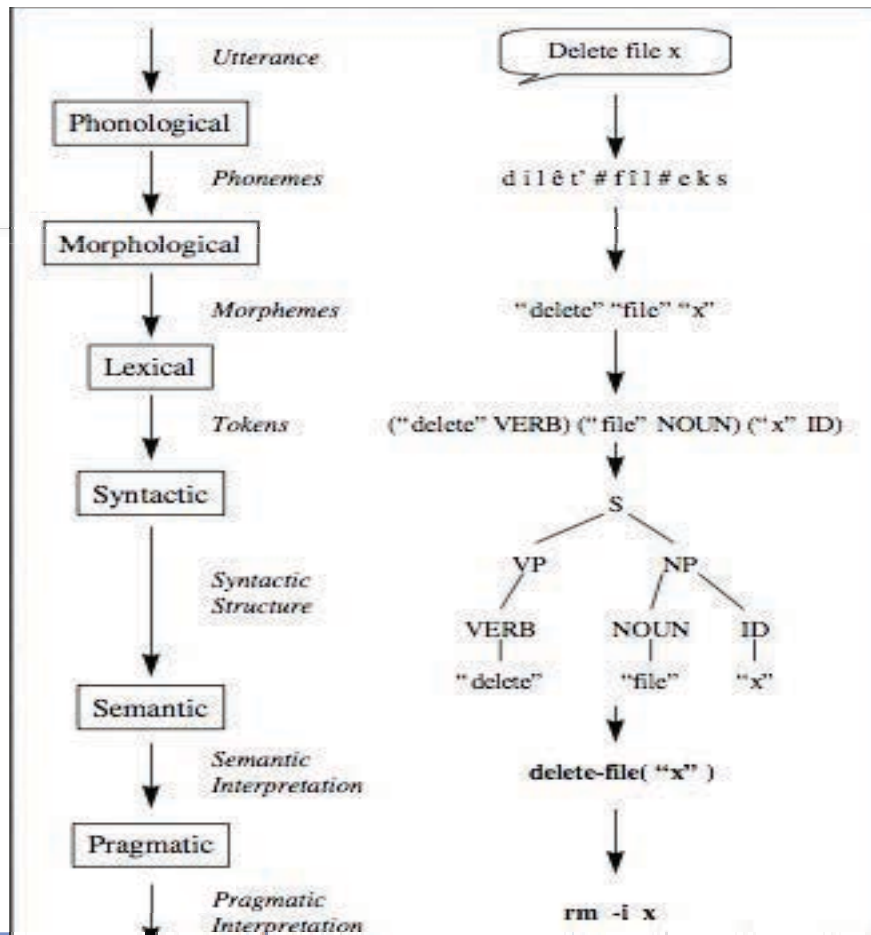
Centro de Estudios en Ingeniería de Software
Depto. Cs. Computación e Informática
Universidad de La Frontera
Temuco, Chile

Indice

- Ingeniería de Requerimientos (RE)
- Procesamiento de Lenguaje Natural (NLP)
- **Avances en la relación entre RE & NLP**
- **Contribuciones Potenciales**
- **Desafíos de Investigación**

Después de este tutorial, podrás :

- Definir RE y su describir su importancia.
- Definir NLP y sus áreas de acción.
- **Conocer el estado del arte en la relación entre NLP y RE.**
- **Describir las contribuciones potenciales de NLP para RE.**
- **Listar los principales desafíos para la investigación en RE.**



RE & NLP

RE main challenges

Sommerville, I.

1. Changing requirements
2. Lack of standardization
3. Differing perspectives
4. People and politics

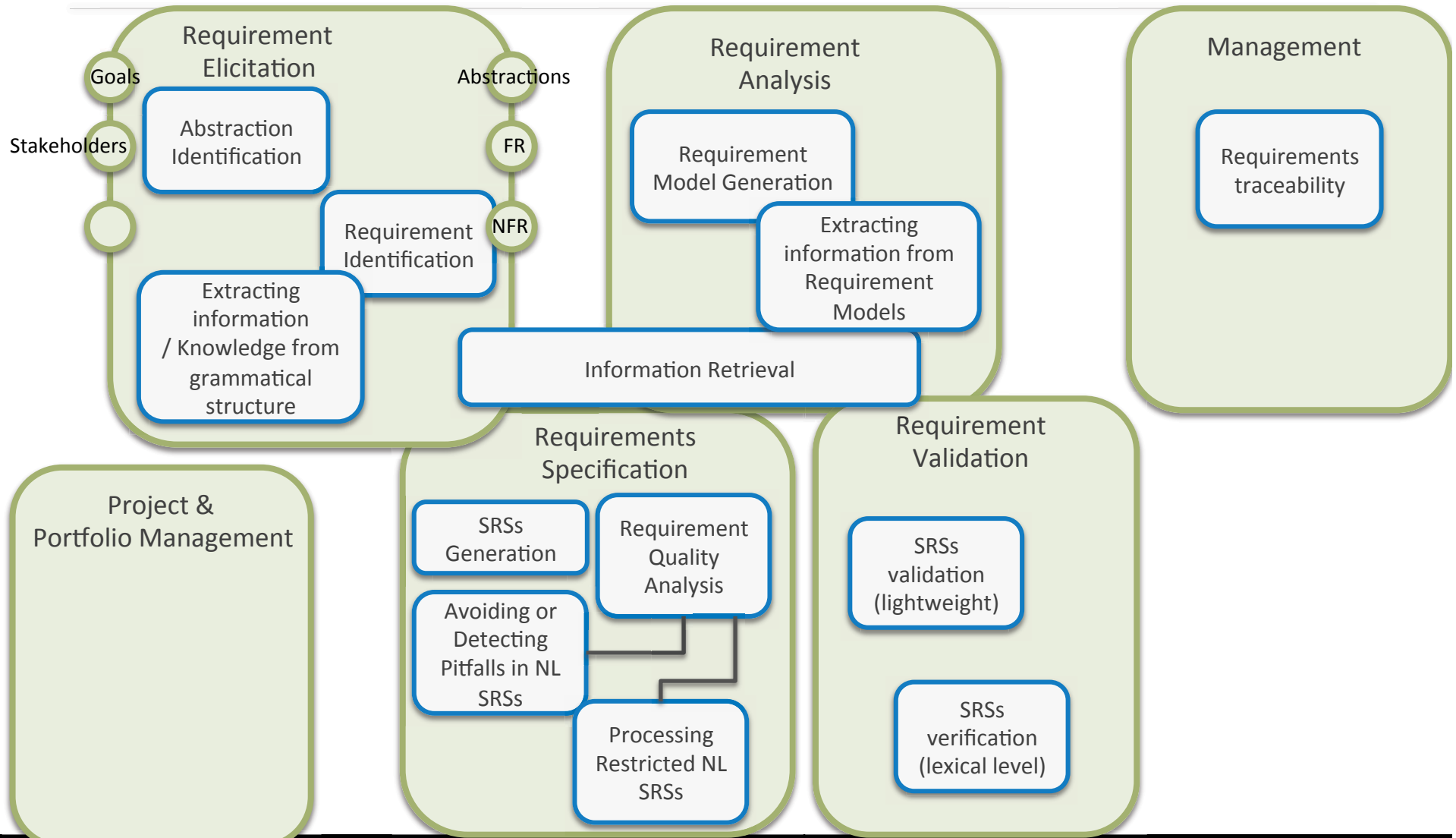
Ian Sommerville. 2015. Software Engineering: 10th Edition)-International Computer Science. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Bashar Nuseibeh and Steve Easterbrook (2000)

1. Development of new techniques for formally modelling and analysing properties of the environment
2. Bridging the gap between requirements elicitation approaches based on contextual enquiry and more formal specification and analysis techniques .
3. Richer models for capturing and analysing non-functional requirements
4. Better understanding of the impact of software architectural choices on the prioritisation and evolution of requirements.
5. Reuse of requirements models.
6. Multidisciplinary training for requirements practitioners.

Bashar Nuseibeh and Steve Easterbrook. 2000. Requirements engineering: a roadmap. In Proceedings of the Conference on The Future of Software Engineering (ICSE '00). ACM, New York, NY, USA, 35-46

NLP in RE – Current Potential Contributions

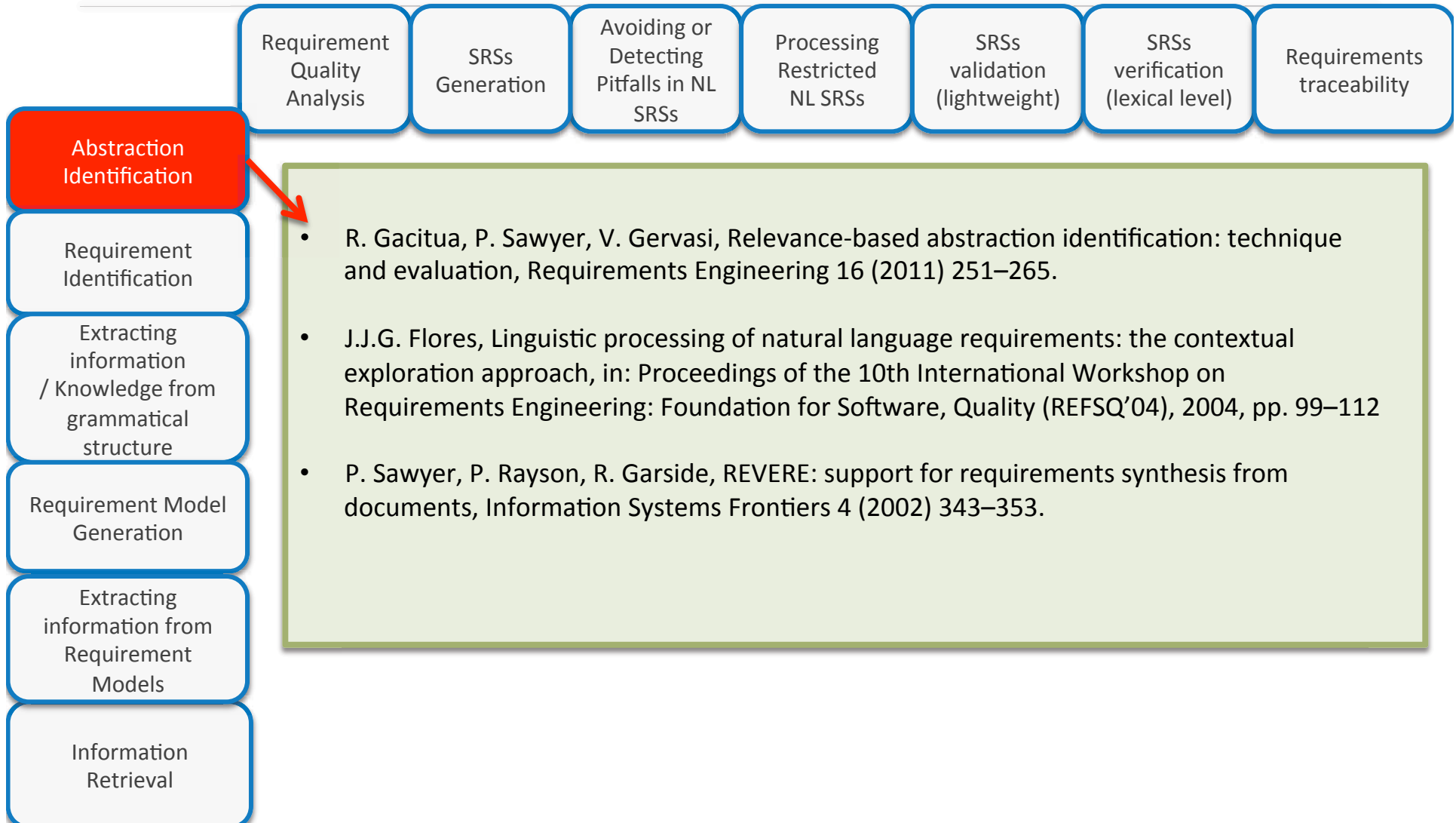




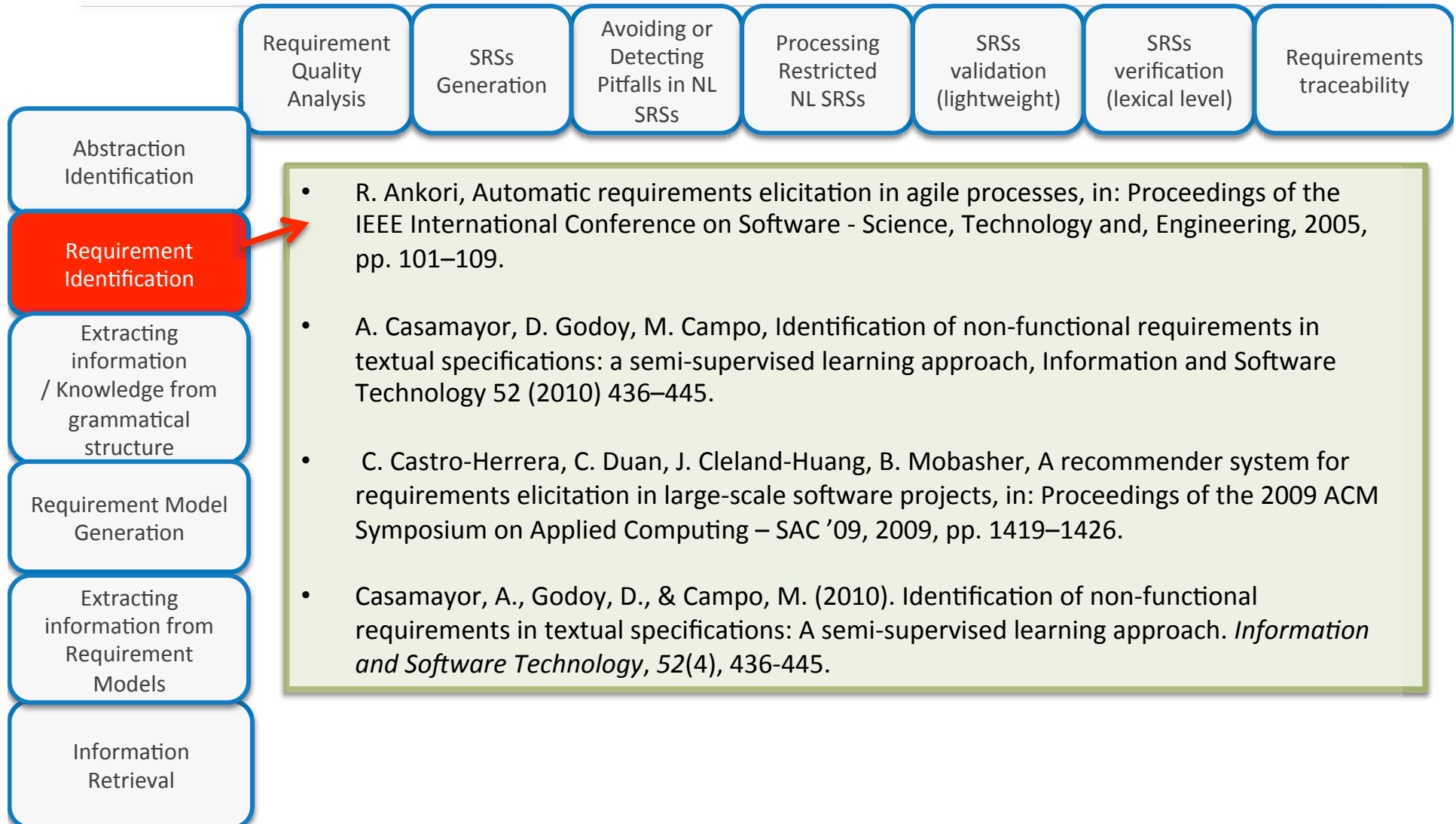
LITERATURE REVIEW

Evidence of Background Information

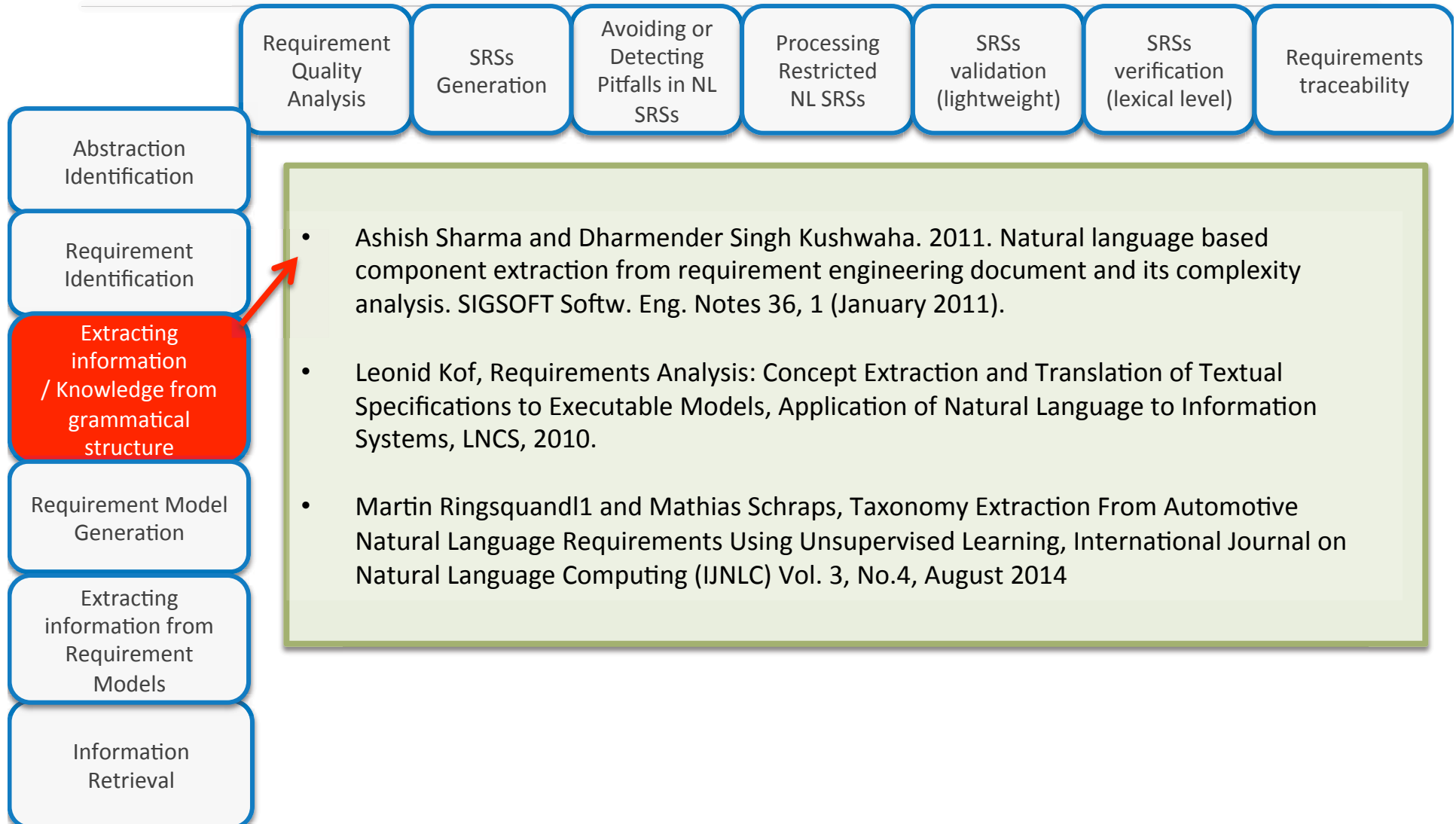
NLP in RE → Abstraction Identification



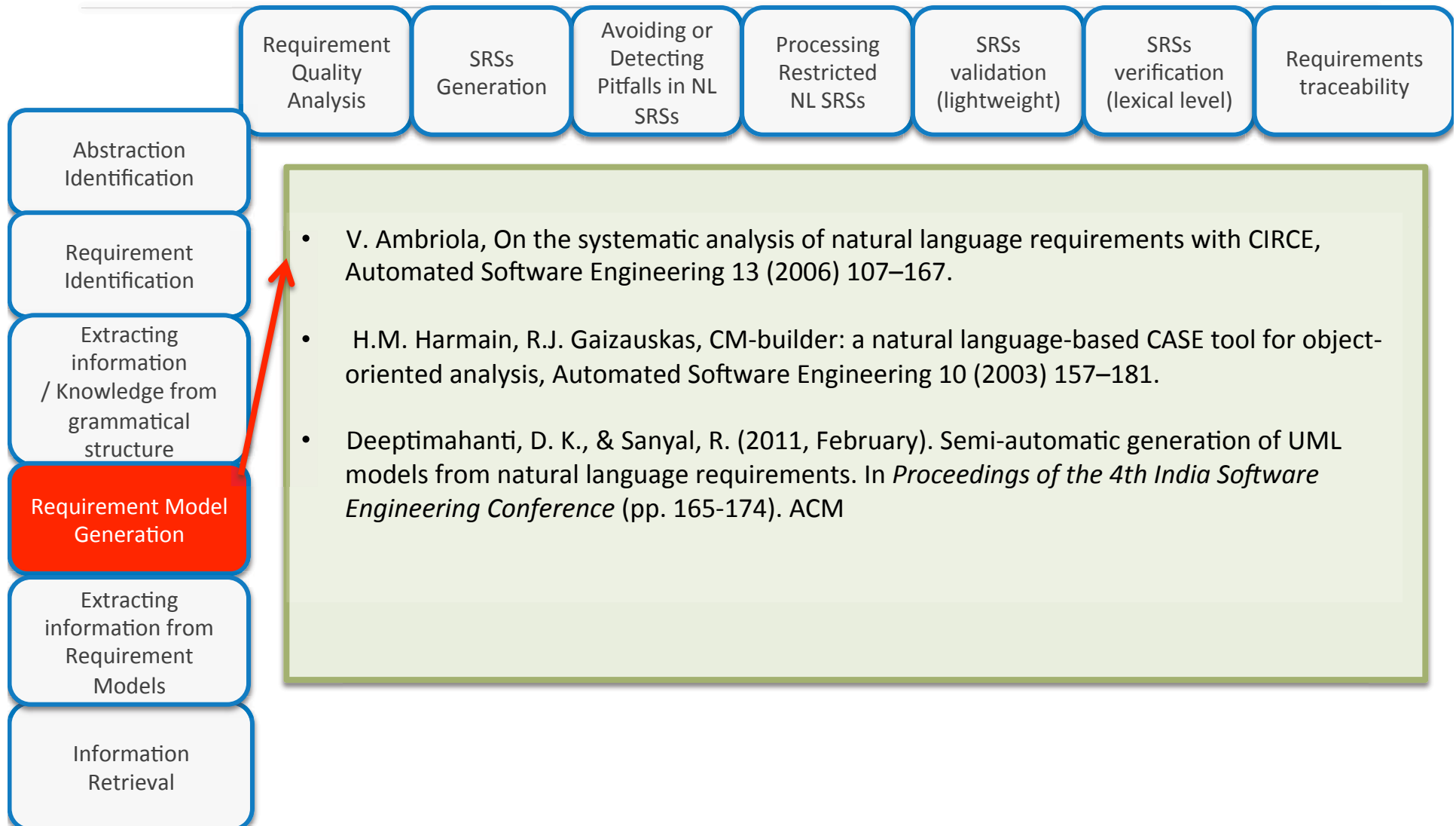
NLP in RE -> Requirement Identification



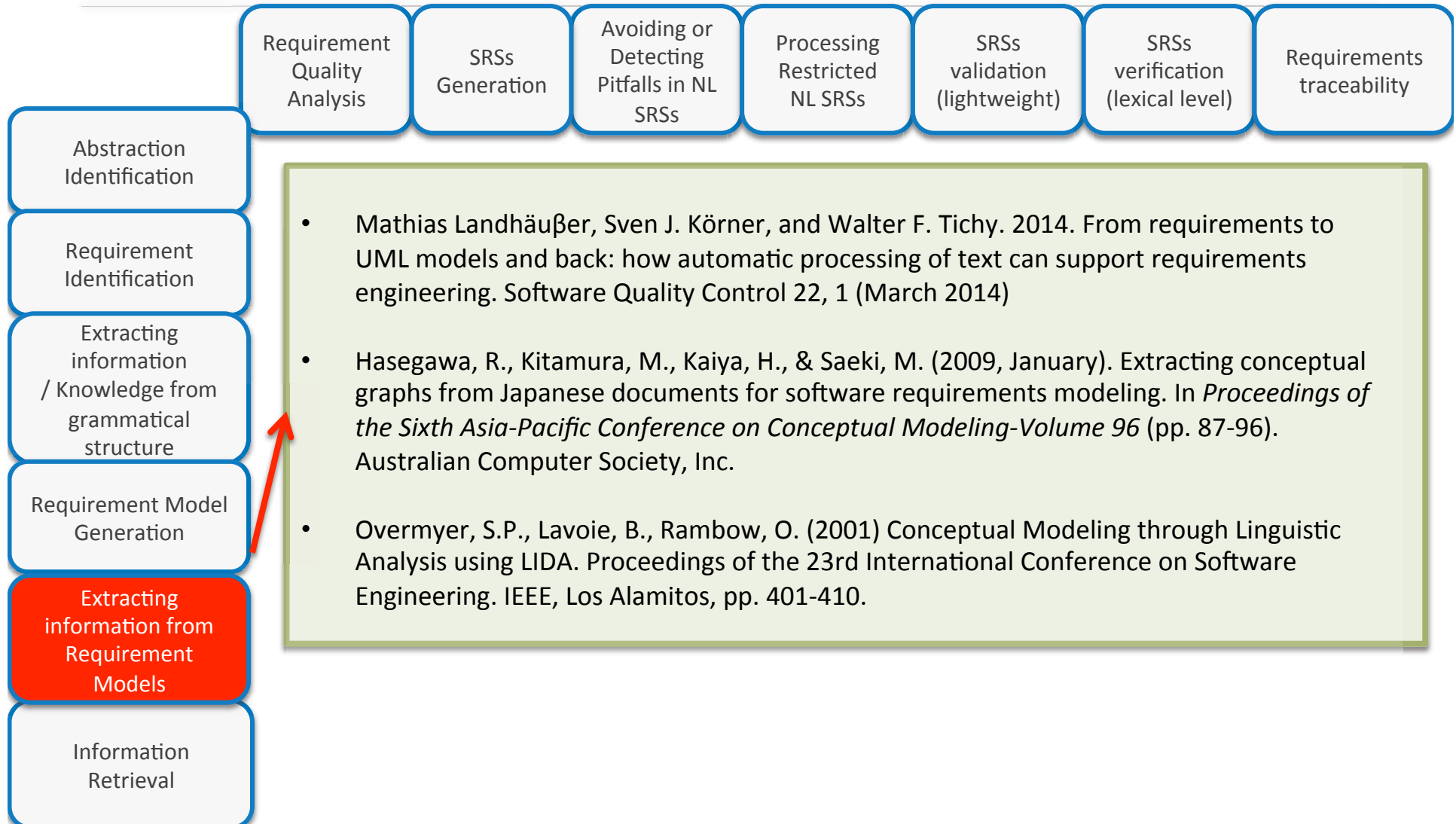
NLP in RE -> Extracting information



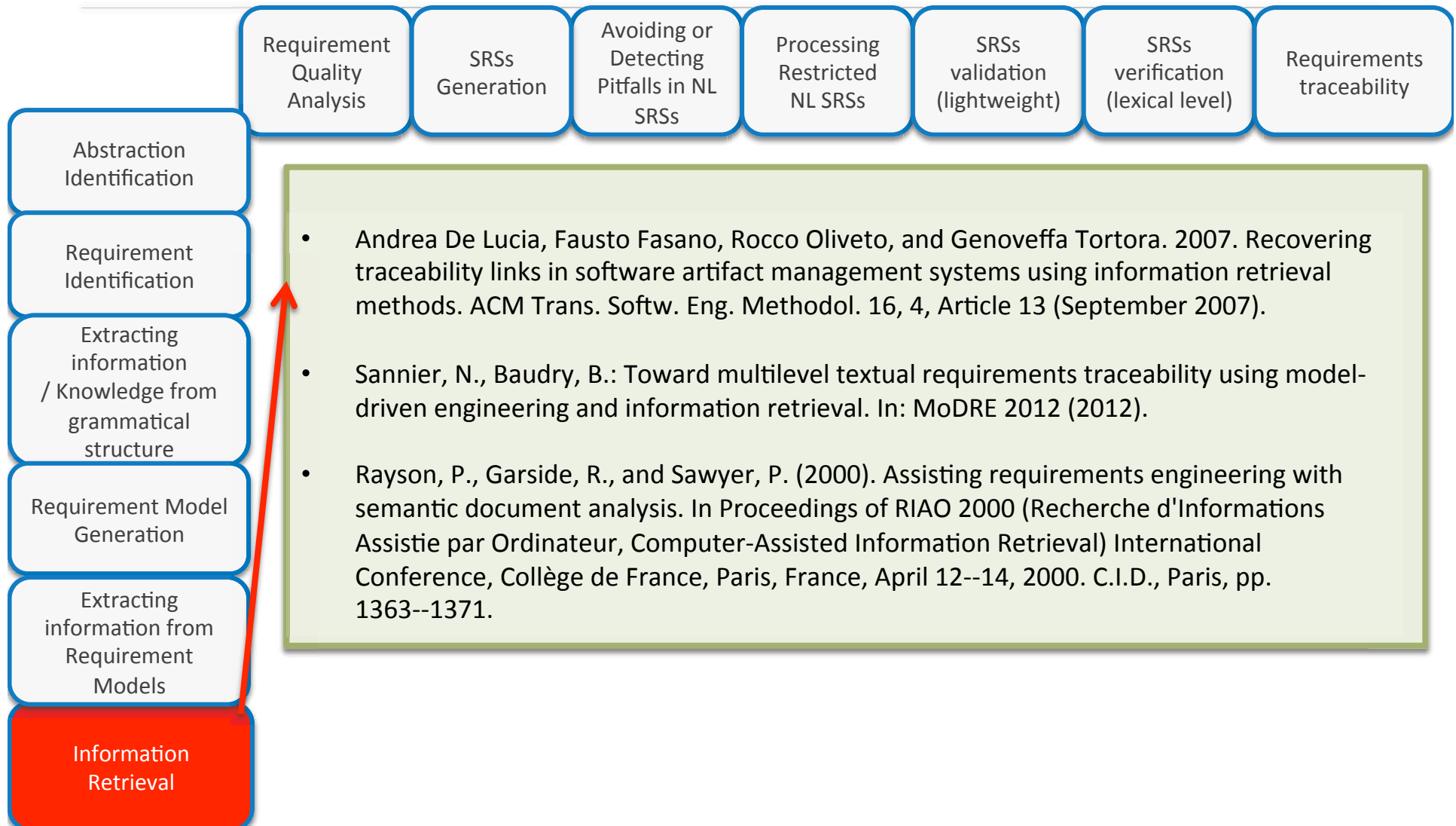
NLP in RE → Requirement Model Generation



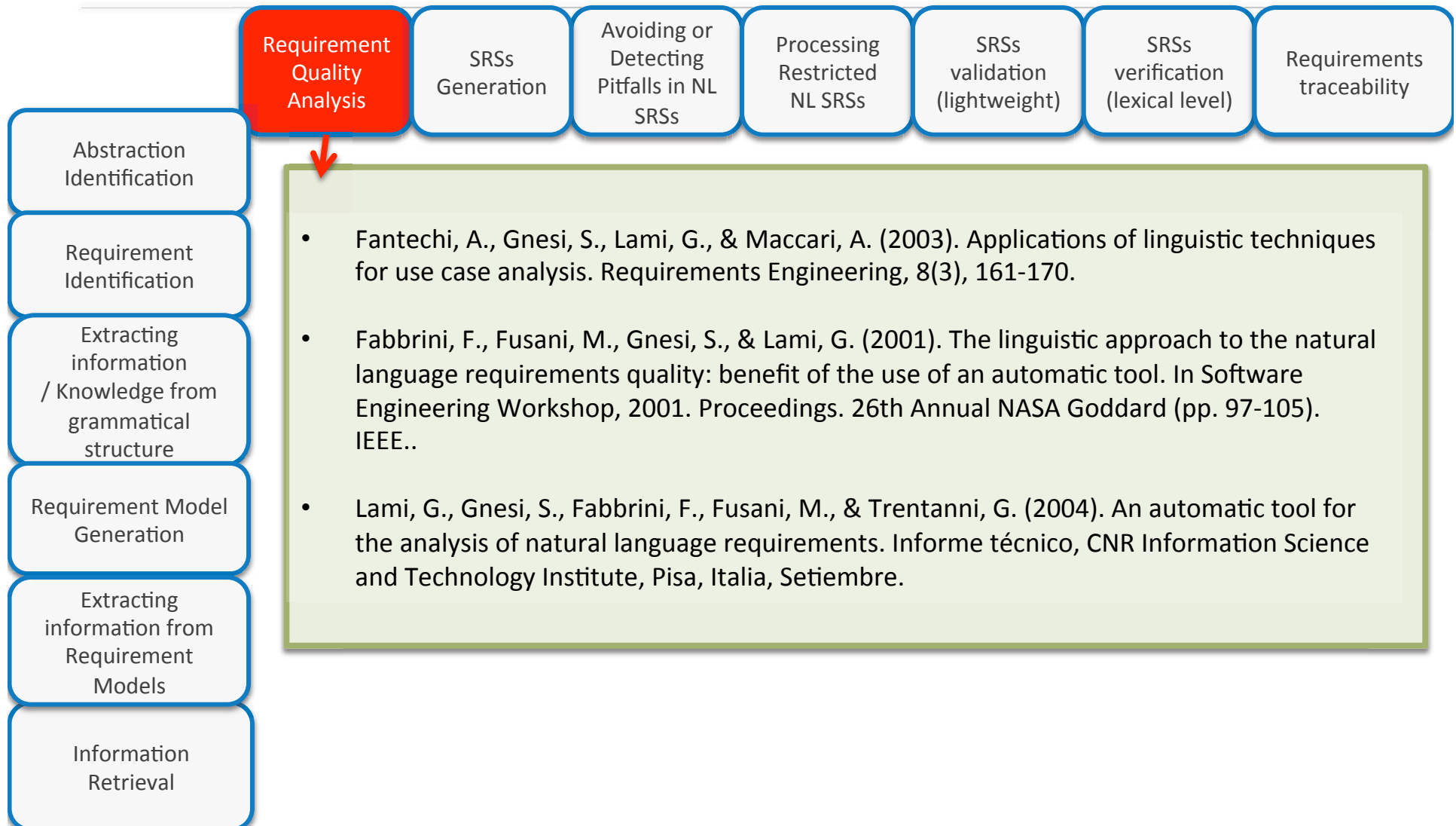
NLP in RE ->Extracting information from Req. Models



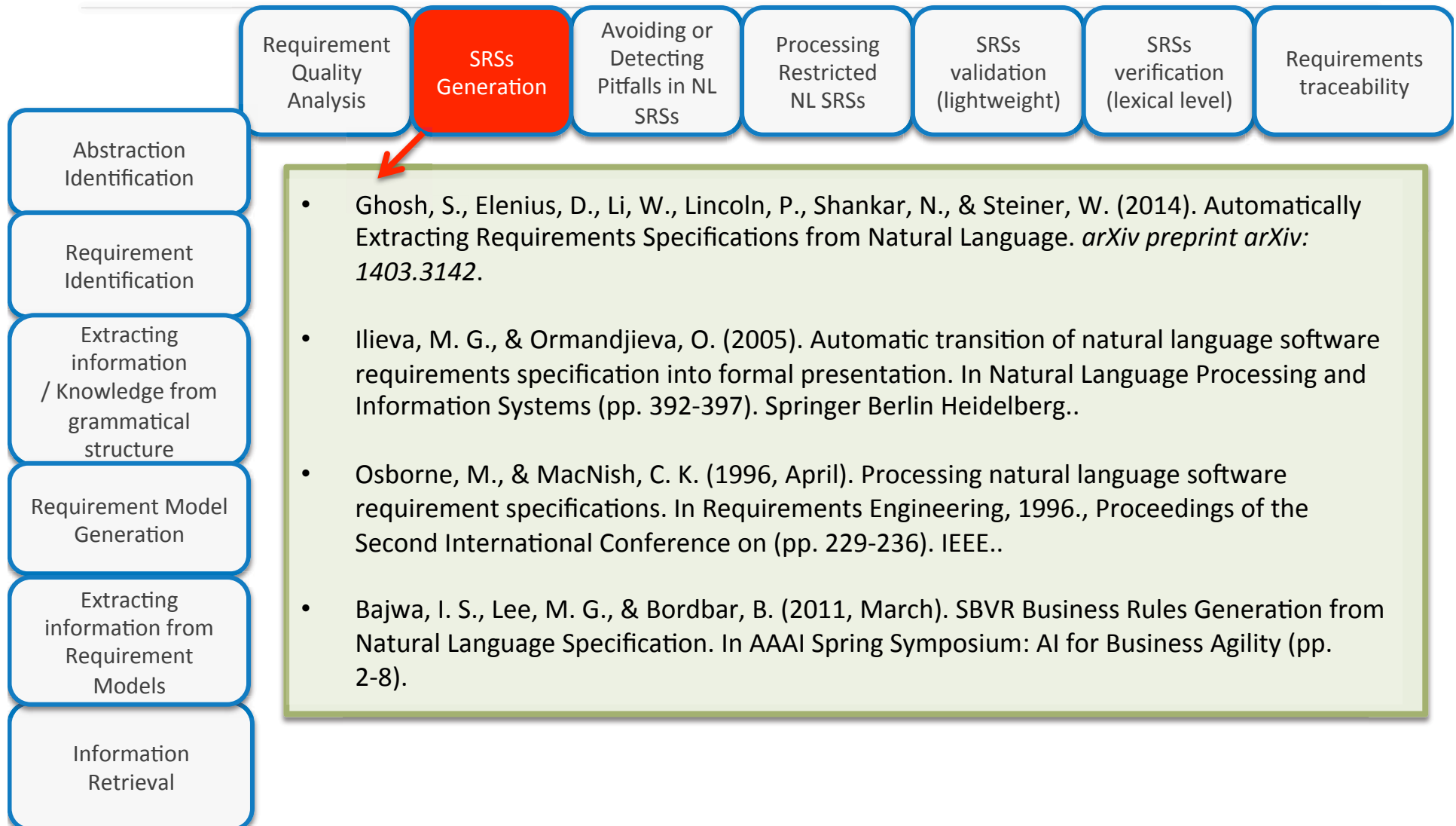
NLP in RE -> Information Retrieval



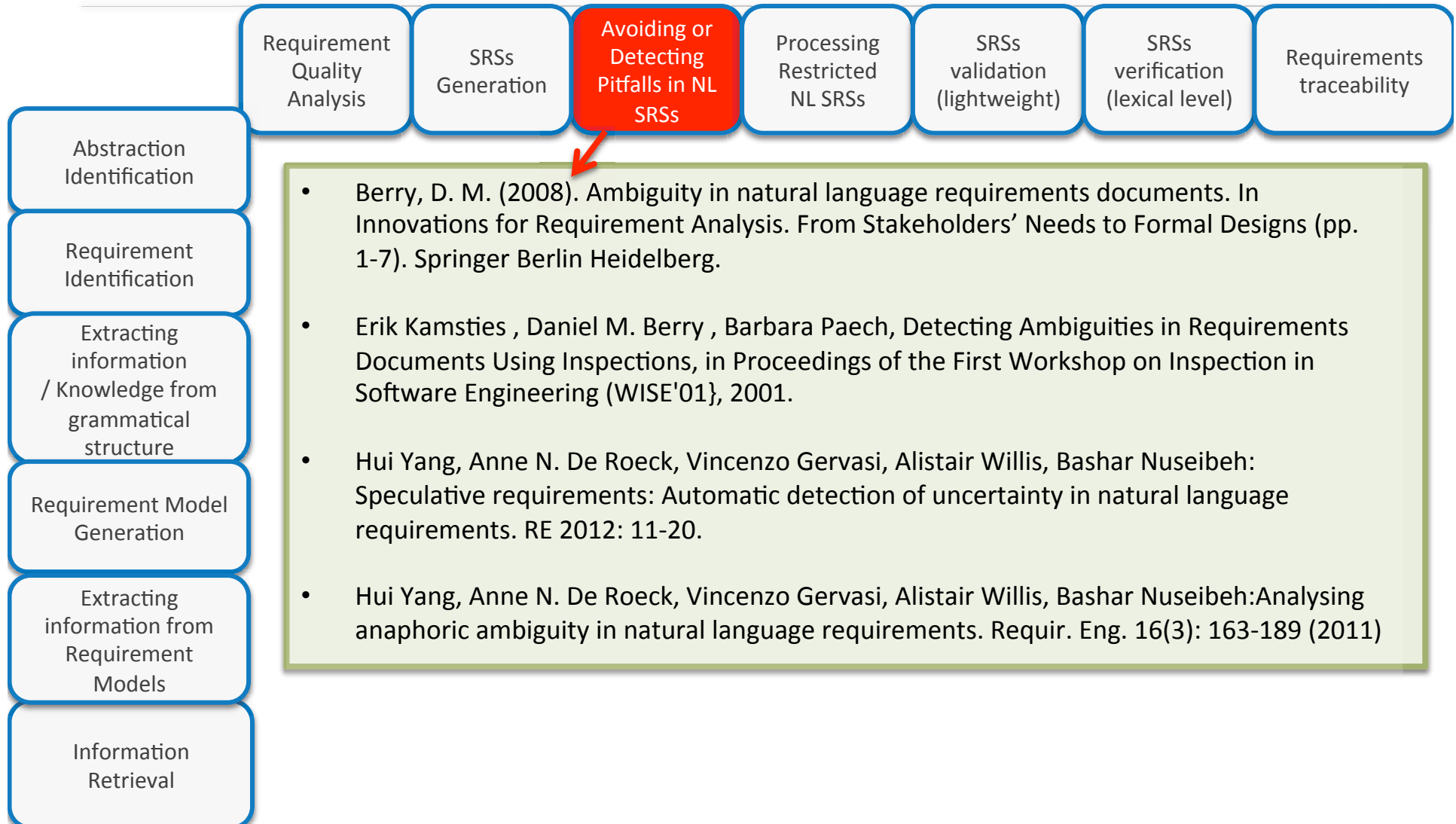
NLP in RE -> Requirements Quality Analysis



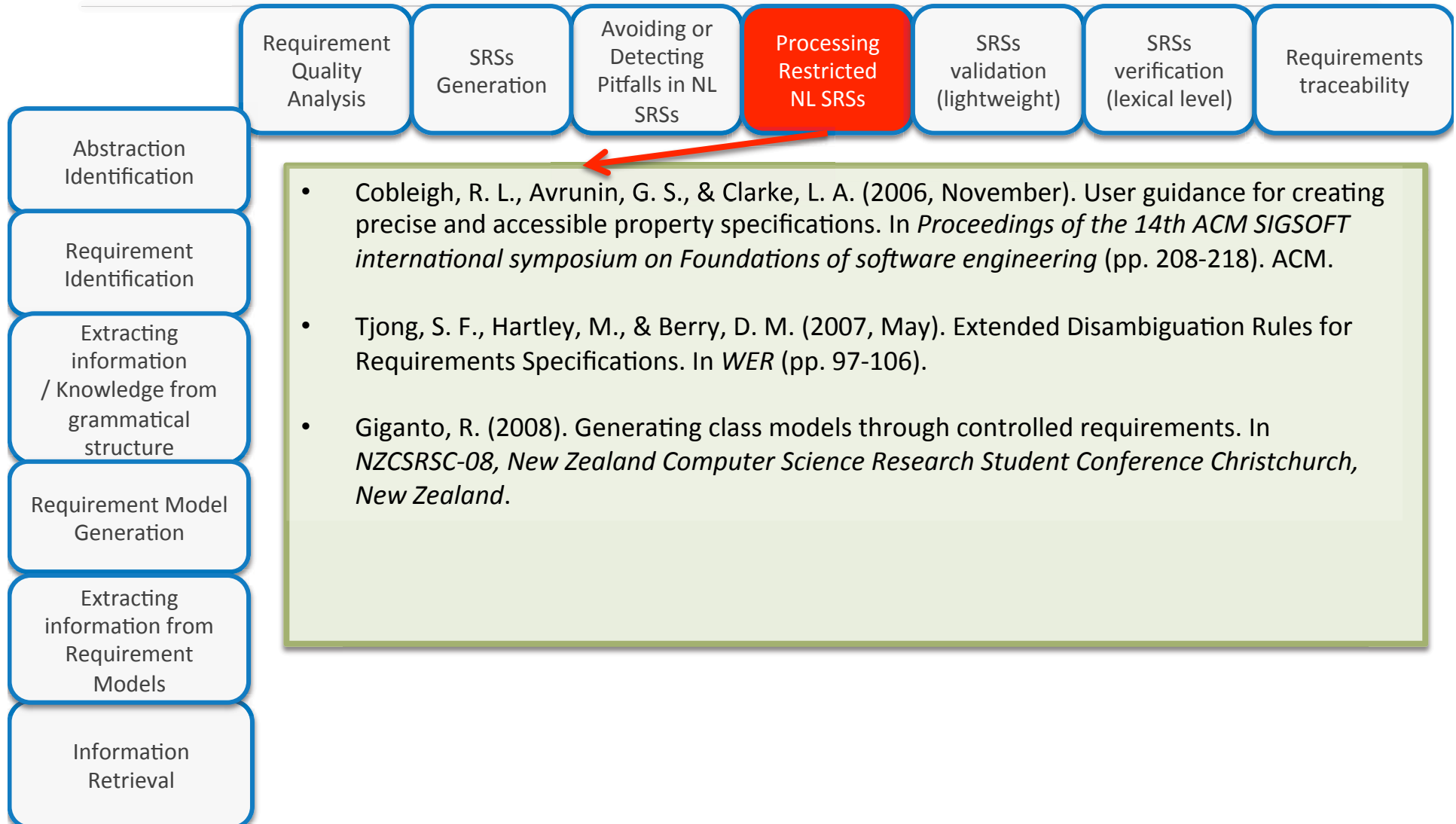
NLP in RE -> SRSs Generation



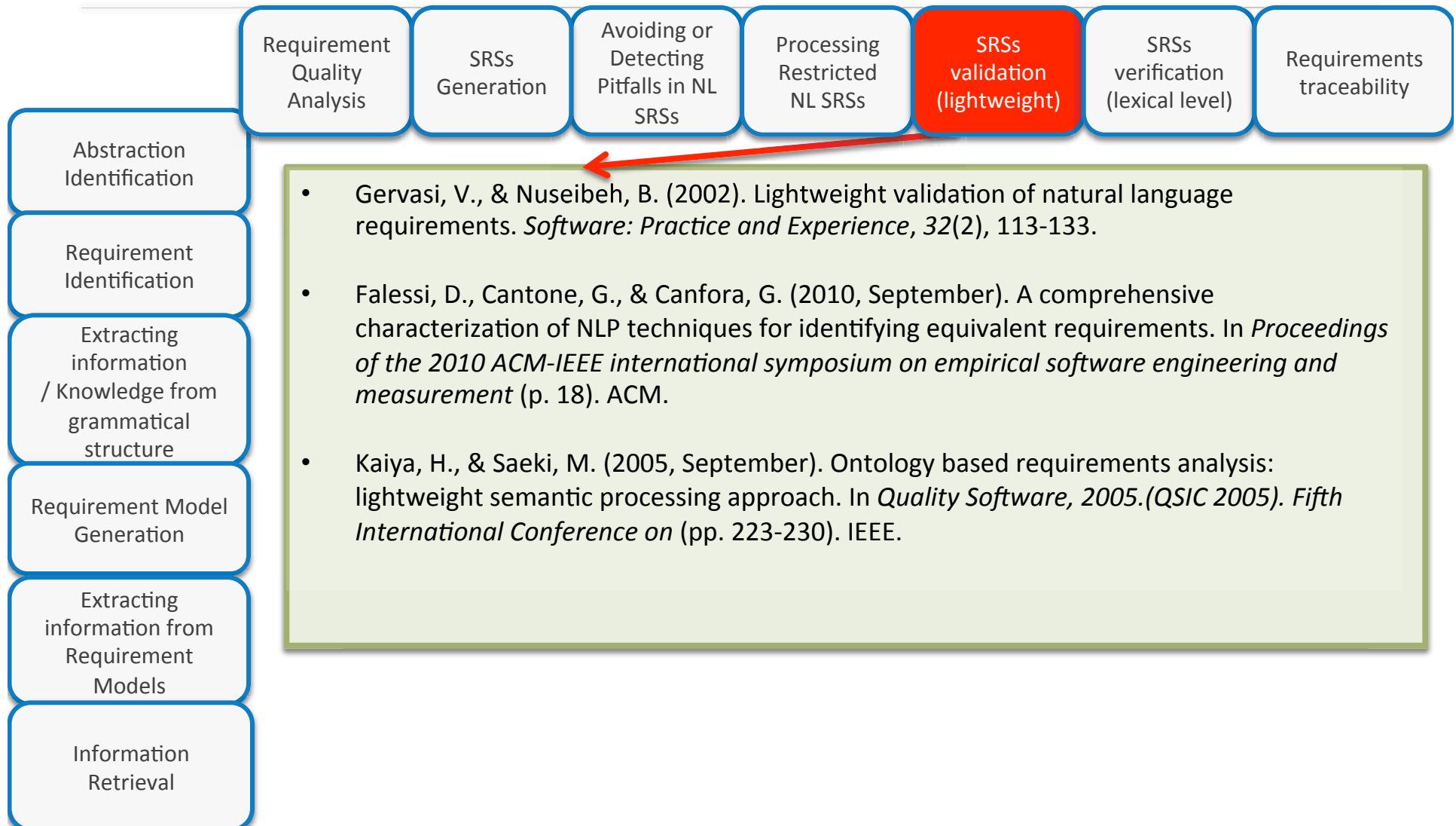
NLP in RE -> Avoiding or Detecting Pitfalls in SRSs



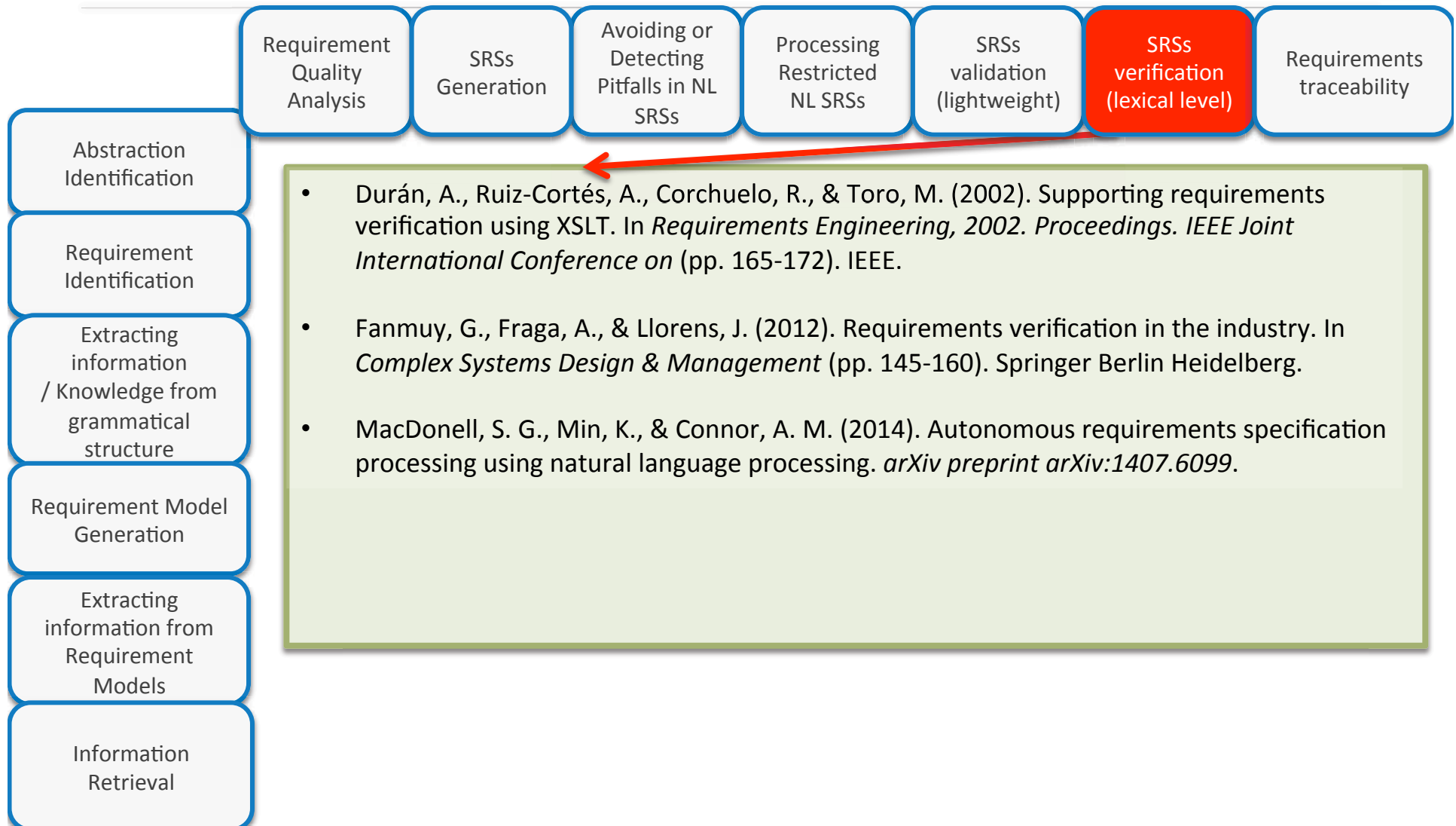
NLP in RE -> Processing Restricted NL SRSs



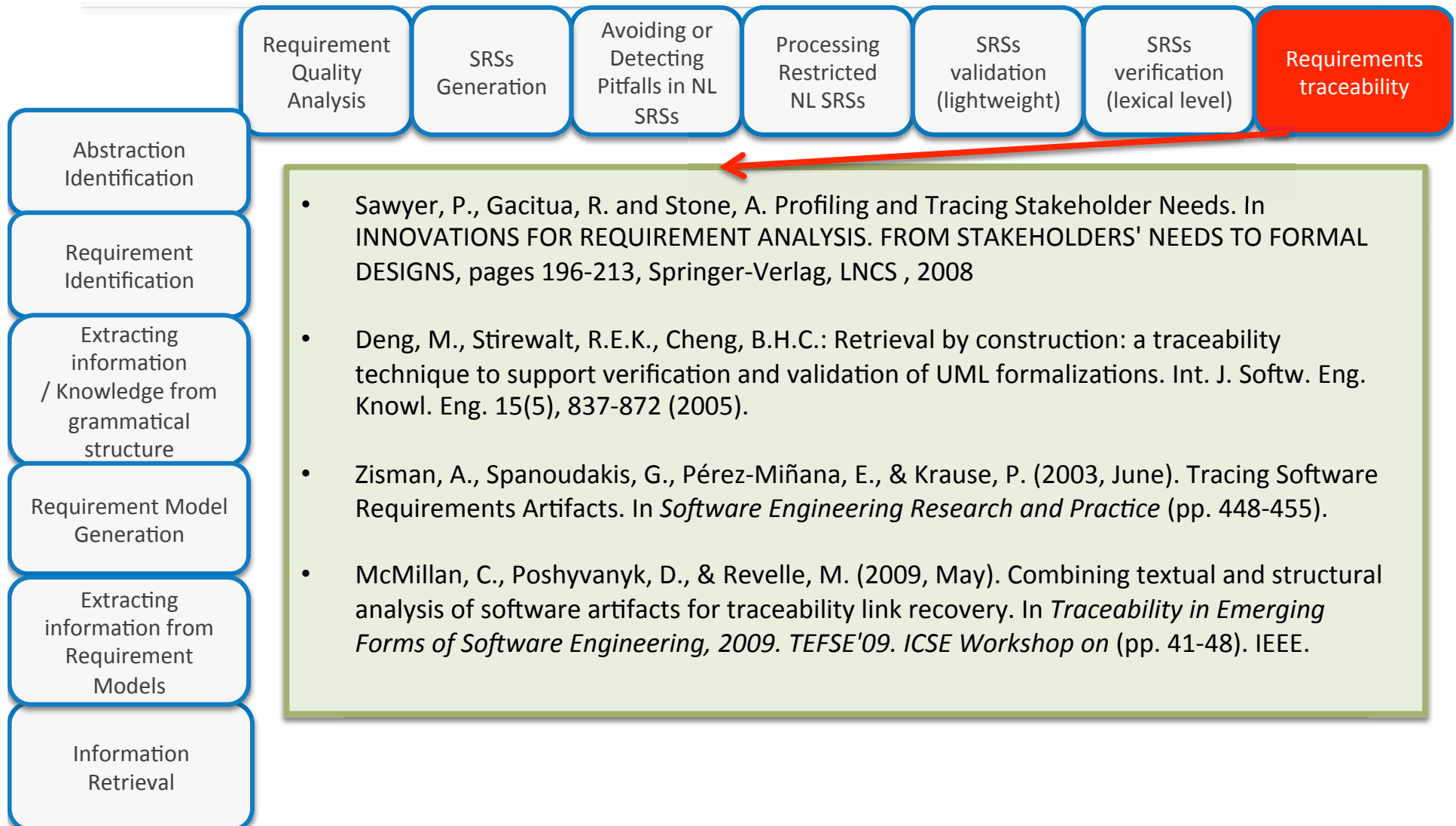
NLP in RE -> SRSs Validation



NLP in RE -> SRSs Verification



NLP in RE -> Requirements traceability



RE & NLP Considerations

- TEXTUAL FORM: Requirements normally does not exist in textual form.
 - Some information occurs naturally as text, typically process descriptions or predefined procedures, but much more is to be found in diagrams or in the physical reality surrounding the client.
 - Assuming however that the requirements definition task is being performed by an intelligent human and that a substantial body of machine readable text is available, there is no doubt that tools to scan, search, browse and tag that text could assist in developing a full and accurate statement of needs.
- COMPLEXITY:
 - Understanding in *NLP* seems to be a fractal-like problem.
 - Narrow domain understanding of natural language may be achieved in the medium term.
- COGNITIVE AND SOCIAL:
 - It is not possible to depend on NLP to solve the RE bottle-neck, since systems are, and will increasingly be recognised to be, social organisms, embodying everything from the deterministic microchip to the emotional and personal needs of the people involved.



Tutorial

RESEARCH CHALLENGES

RE Summary

1. **SOCIAL SCIENCES:** RE draws on the cognitive and social sciences:
 - Cognitive Psychology, Anthropology, Sociology, Linguistics
2. **BOUNDARIES:** The identification of stakeholders and user classes, of goals and tasks, and of scenarios and use cases all depend on how the boundaries are chosen.
3. **MODELS:** Models can be used to represent a whole range of products of the RE process.
4. **COMMUNICATION:** RE is not only a process of discovering and specifying requirements, it is also a process of facilitating effective communication of these requirements among different stakeholders.
5. **DOCUMENTATION:** The way in which requirements are documented plays an important role in ensuring that they can be read, analysed, (re-)written, and validated.
6. **VALIDATION:** Requirements validation is difficult for two reasons.
 - The first reason is philosophical in nature, and concerns the question of truth and what is knowable.
 - The second reason is social, and concerns the difficulty of reaching agreement among different stakeholders with conflicting goals.
7. **CHANGES:** Managing changing requirements is not only a process of managing documentation, it is also a process of recognising change through continued requirements elicitation, re- evaluation of risk, and evaluation of systems in their operational

RE Challenges

Sommerville, I.

1. Changing requirements
2. Lack of standardization
3. Differing perspectives
4. People and politics

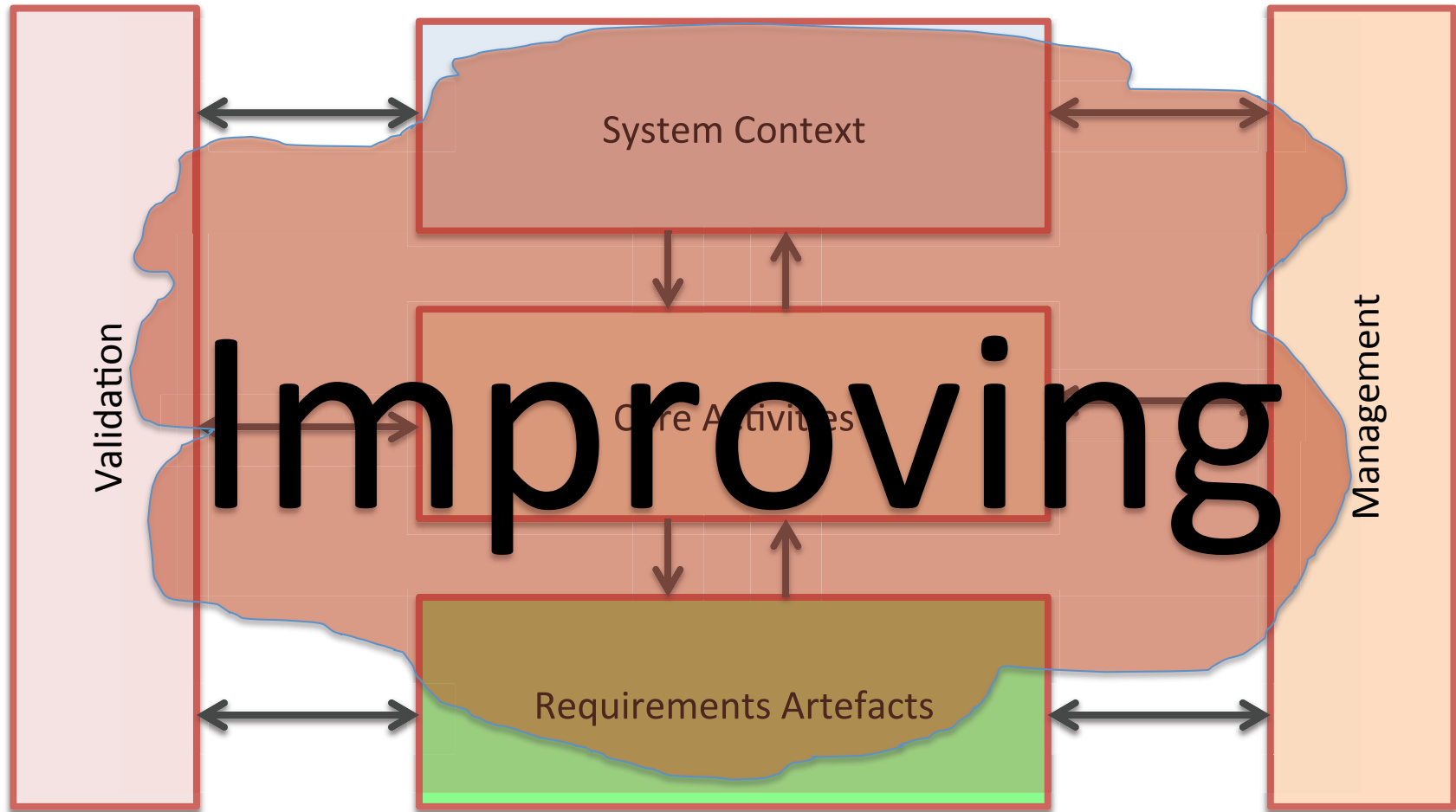
Bashar Nuseibeh and Steve Easterbrook (2000)

1. Development of new techniques for formally modelling and analysing properties of the environment
2. Bridging the gap between requirements elicitation approaches based on contextual enquiry and more formal specification and analysis techniques .
3. Richer models for capturing and analysing non-functional requirements
4. Better understanding of the impact of software architectural choices on the prioritisation and evolution of requirements.
5. Reuse of requirements models.
6. Multidisciplinary training for requirements practitioners.

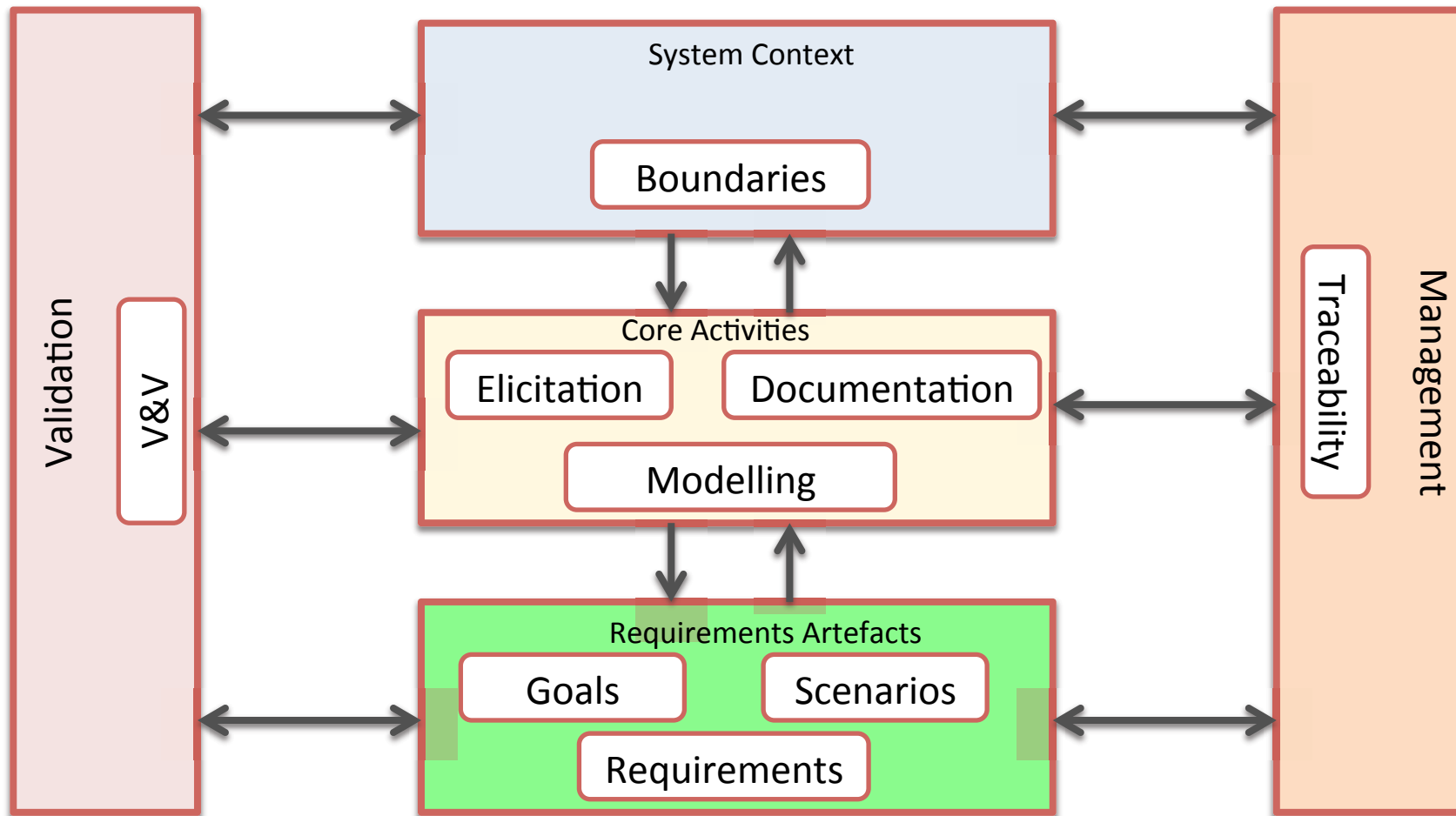
Ian Sommerville. 2015. Software Engineering: 10th Edition)-International Computer Science. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Bashar Nuseibeh and Steve Easterbrook. 2000. Requirements engineering: a roadmap. In Proceedings of the Conference on The Future of Software Engineering (ICSE '00). ACM, New York, NY, USA, 35-46

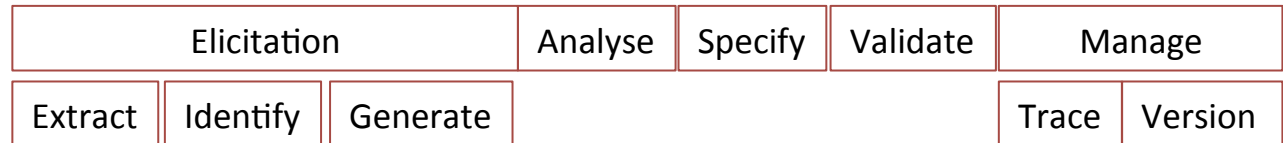
Research Challenges - Framework



Some Research Challenges for NLP in RE



RE & NLP -> Research Challenges



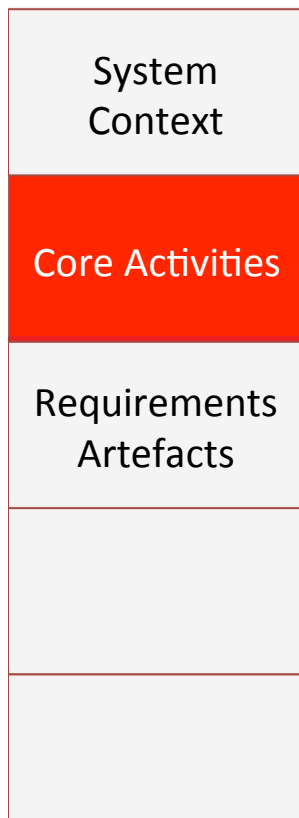
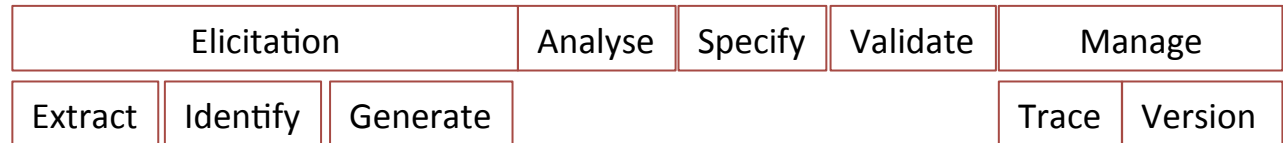
System
Context

Core Activities

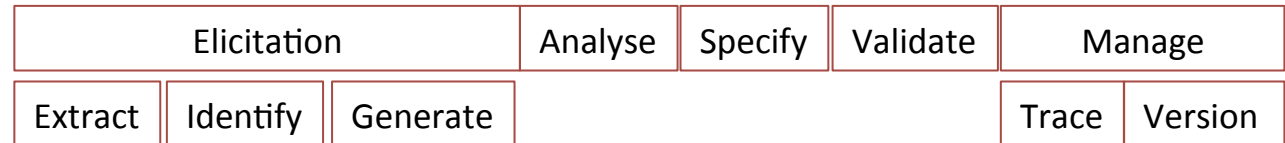
Requirements
Artefacts

1. From speech to text: Automatic interview transcription
2. Automatic extraction and selection of pertinent domain documents.
 - Big Data, Semantic Web
3. Automatic extraction of scenarios
4. Automatic adequacy to standards and guidelines
5. Automatic identification of domain artefacts
 - Domain terminology and Semantic models (e.g. Ontologies)

RE & NLP -> Research Challenges



1. Automatic identification of requirements artefacts
 - FR, NFR, Goals, Stakeholders, Business rules, constraints etc.
2. Automatic synthesize and validation of domain models
3. Automatic validation of Requirements
 - Consistency, coherence, duplicity, etc.
4. Automatic improvement of requirements
 - Detection of ambiguity,
5. Automatic validation of SRSs
6. Automatic building SRSs from domain models and NL text

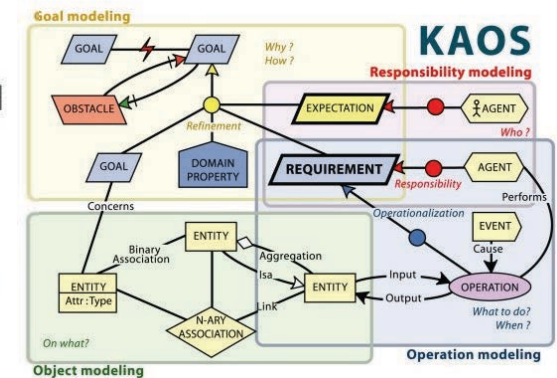
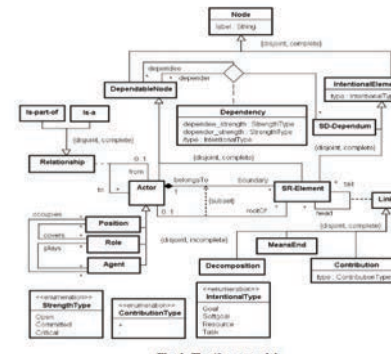


System
Context

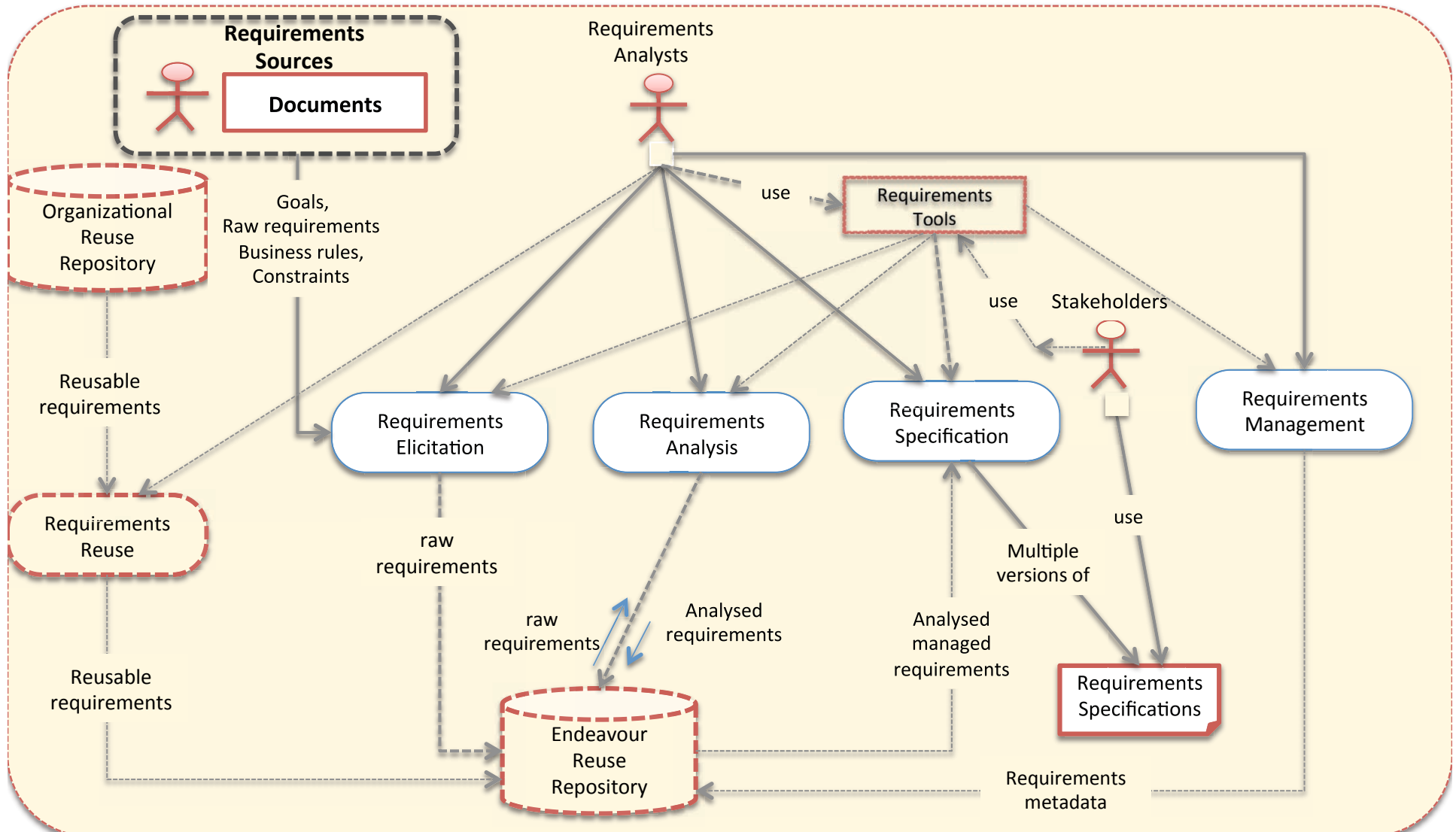
Core Activities

Requirements
Artefacts

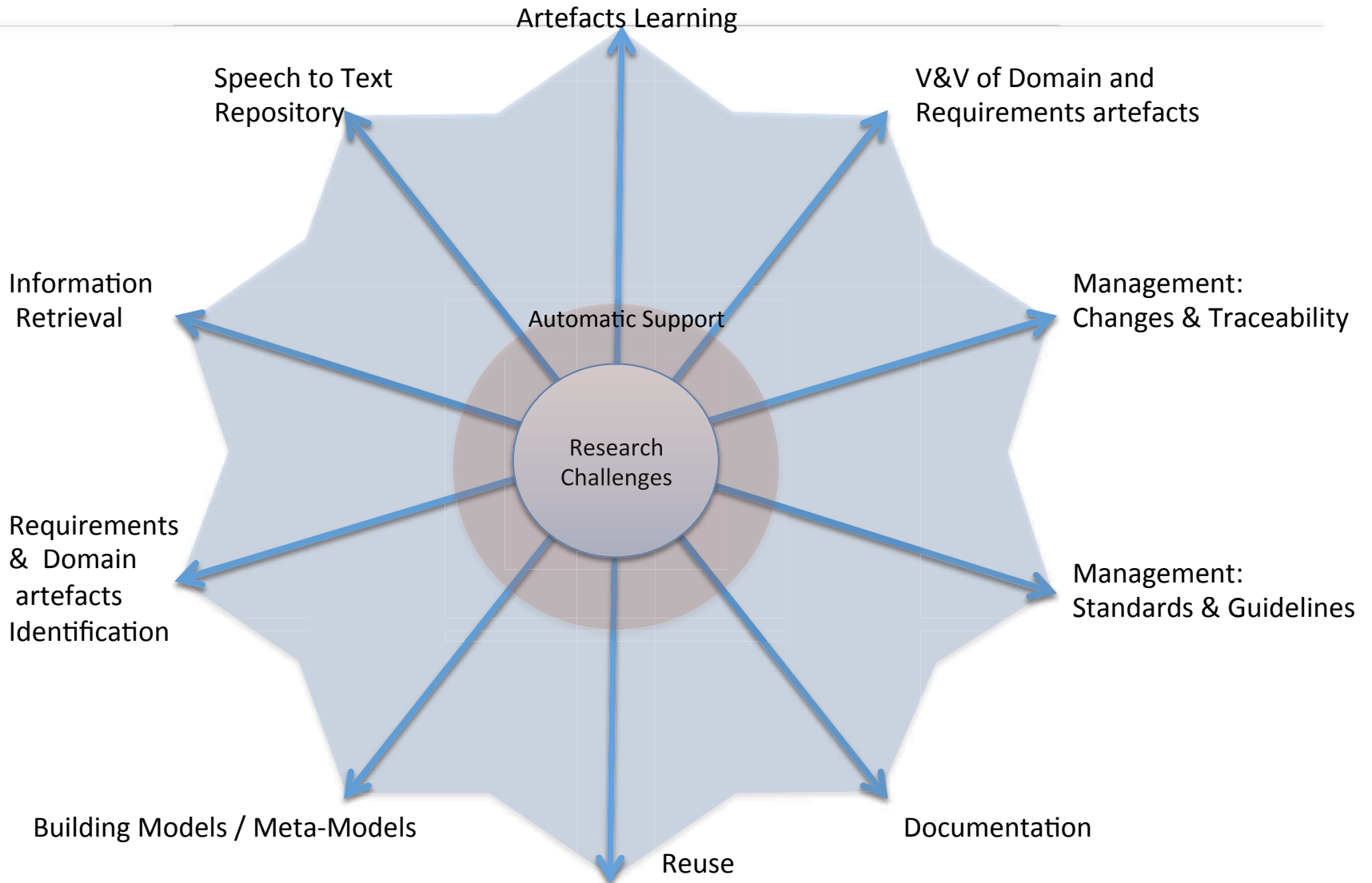
1. Reuse of domain and requirements artefacts
2. Scenarios Learning
3. Automatic building models of requirements artefacts from Meta-Models
 - Example: i* Metamodel and Kaos Metamodel



RE & NLP → Research Challenges → Scope



RE & NLP -> Some Research Challenges



Conclusion

- Requirements engineering is a non-trivial task and NLP is not able to solve all the requirements engineering problems.
 - However, it tackles an extremely important step, namely providing preliminary requirements and domain artefacts (e.g. preliminary models, preliminary lists of requirements and document summaries) for the stakeholders and requirements engineers.
- It is in supporting the social process, and not in supplanting it, that natural language processing will have its proper role.

TUTORIAL

THANK YOU

