Utilize Natural Language Processing in Evidence-based software engineering

Speaker：Jinyu Ma
Software institute, Nanjing University
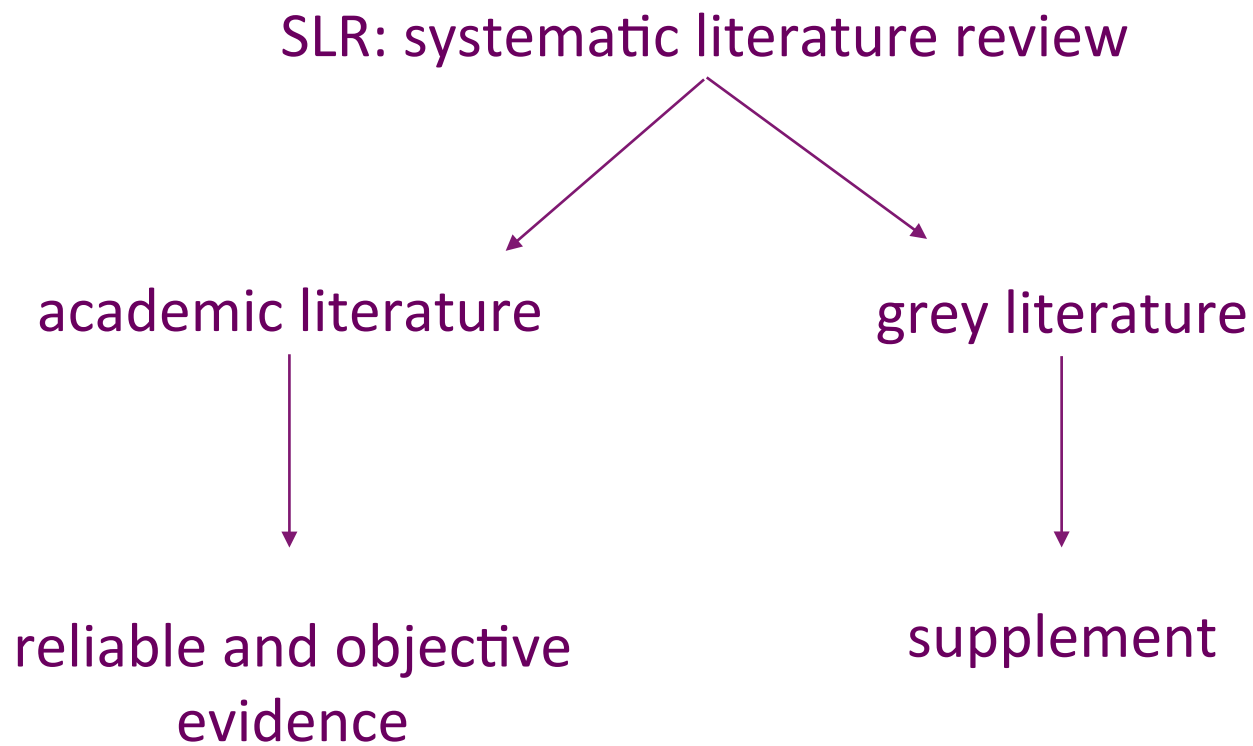2018/6/6

CONTENTS

Background

EBSE & SLR

> **Evidence-Based Software Engineering (EBSE):** systematically answer particular research questions by identifying, evaluating and synthesizing relevant evidence.

> **Systematic Literature Review (SLR):** Kitchenham et al. referred to the studies in Medicine and proposed a guideline for SLR in software engineering in 2004.

Background

SLR: systematic literature review

academic literature

grey literature

reliable and objective evidence

supplement

Background

Grey Literature Review

- The strongest feature of grey literature : large amount

- The guideline for traditional SLR is no longer suitable for grey literature.

How to automatically perform grey literature review?

NANJING UNIVERSITY

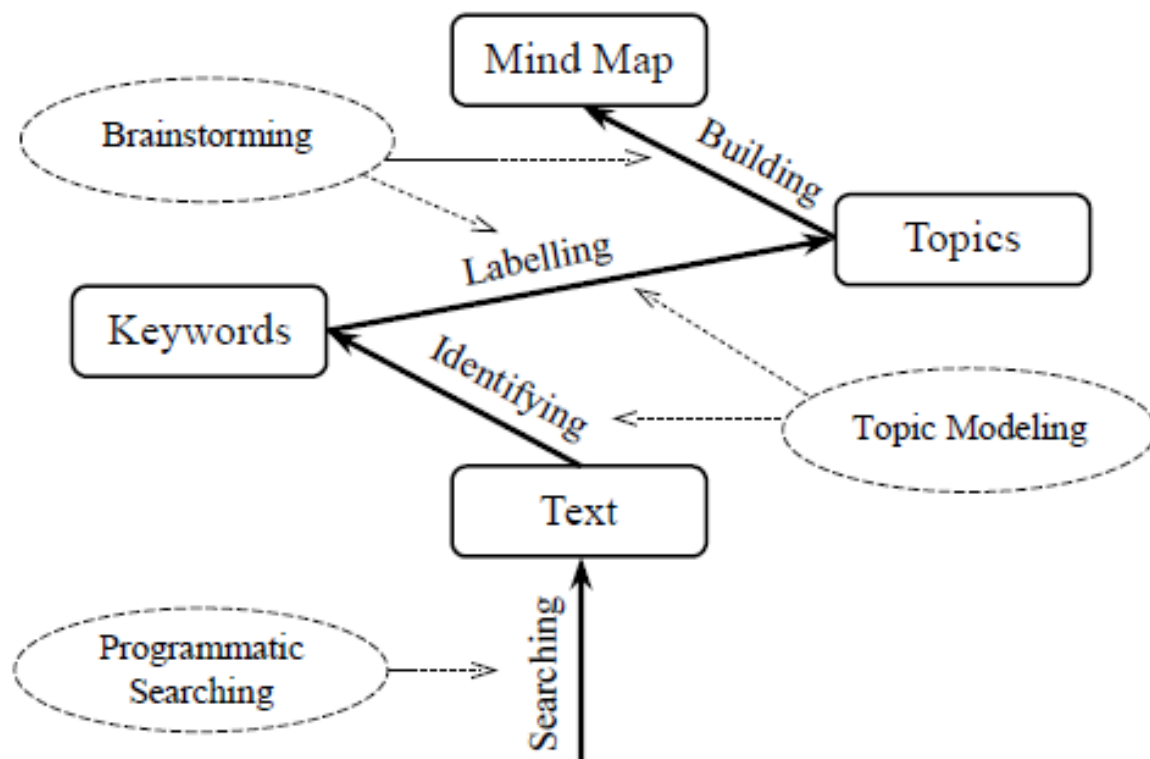# CONTENTS

Topic Modeling

Background

> ➢ Topic modeling is a statistical model that aims to abstract latent topics from a series of documents.

> ➢ It believes that a document contains several topics with different probabilities, and topics are made up with baskets of words with different probabilities

$$P(word|doc) = \sum_{topic} P(word|topic) \times P(topic|doc)$$
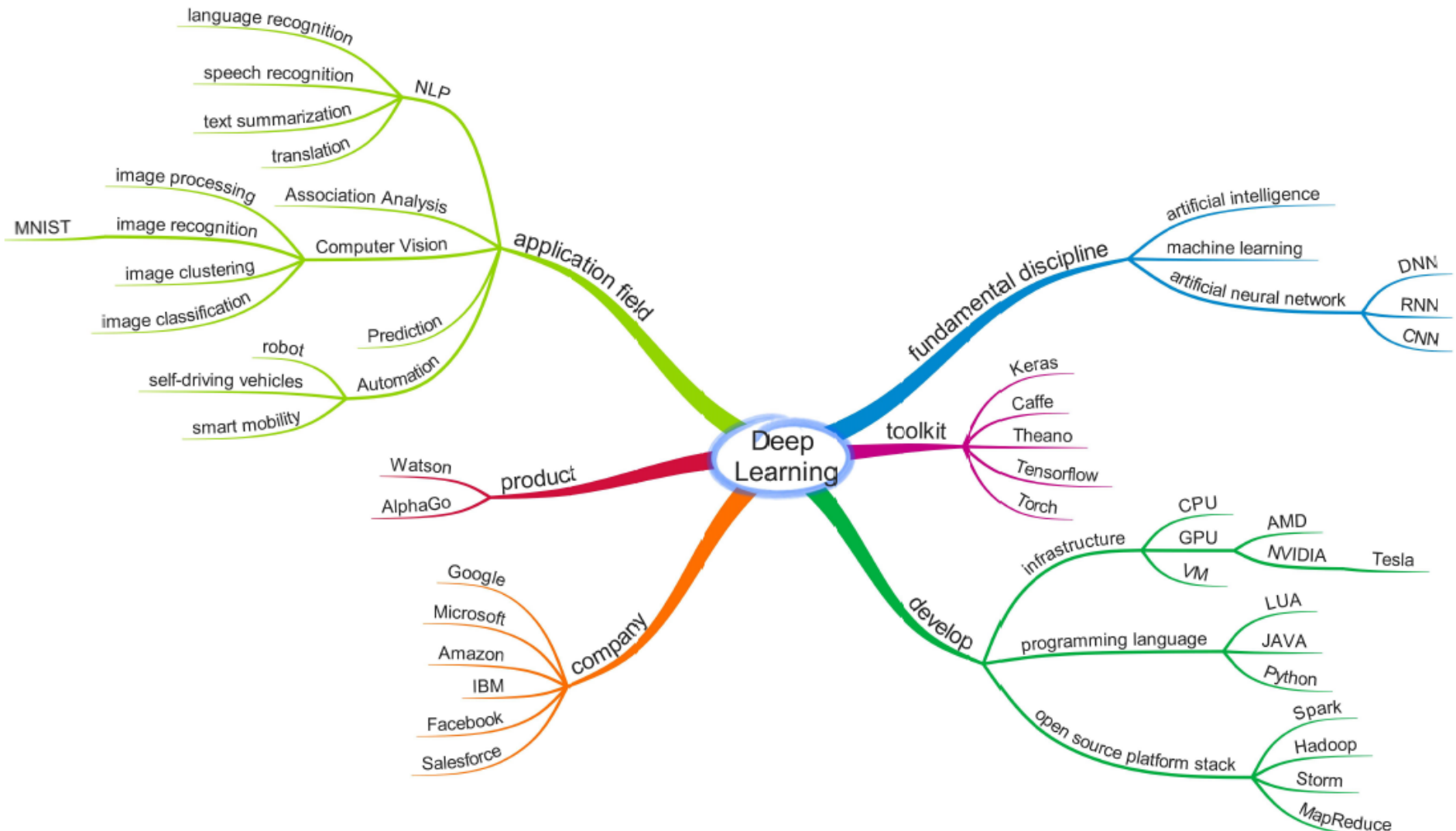
NANJING UNIVERSITY

Topic Modeling

Topic Modeling

Method

> **Tool :** MALLET

> **Research source:**

1. Technology discussions in a Web community.   ——Stack Overflow

2. Blogs/news from an IT website. ——ZDNet

3. Distilled scholarly discourse in an accessible forum.

    ——Tiny Transactions on Computer Science —> abstracts in IEEE

4. Invention descriptions in an open-access patent database.

    ——United States Patent and Trademark Office (USPTO) Portal

Topic Modeling

Deep Learning

application field
- NLP
  - language recognition
  - speech recognition
  - text summarization
  - translation
- Association Analysis
- Computer Vision
  - image processing
  - image recognition
    - MNIST
  - image clustering
  - image classification
- Prediction
- Automation
  - robot
  - self-driving vehicles
  - smart mobility

product
- Watson
- AlphaGo

company
- Google
- Microsoft
- Amazon
- IBM
- Facebook
- Salesforce

fundamental discipline
- artificial intelligence
- machine learning
- artificial neural network
  - DNN
  - RNN
  - CNN

toolkit
- Keras
- Caffe
- Theano
- Tensorflow
- Torch

develop
- infrastructure
  - CPU
  - GPU
    - AMD
    - NVIDIA
      - Tesla
  - VM
- programming language
  - LUA
  - JAVA
  - Python
- open source platform stack
  - Spark
  - Hadoop
  - Storm
  - MapReduce

CONTENTS

Current Work——An SLR for NLP

Natural Language Processing

> An area of computer science and artificial intelligence concerned with the interactions between computers and natural languages.

> Deal with various kinds of natural language tasks in different aspects like lexical, syntactic, semantic.

Current Work——An SLR for NLP

Motivation

> To get a comprehensive overview of Natural language Processing

> Focus not only on NLP, but also software engineering.

> To find some useful knowledge for performing automatically grey literature review.

Current Work——An SLR for NLP

Research Questions

> RQ1: What senarios in software engineering can NLP be applied in?

> RQ2: What's the process model of NLP studies in software engineering?

> RQ3: Are there any factors influence the performance of NLP in software engineering?

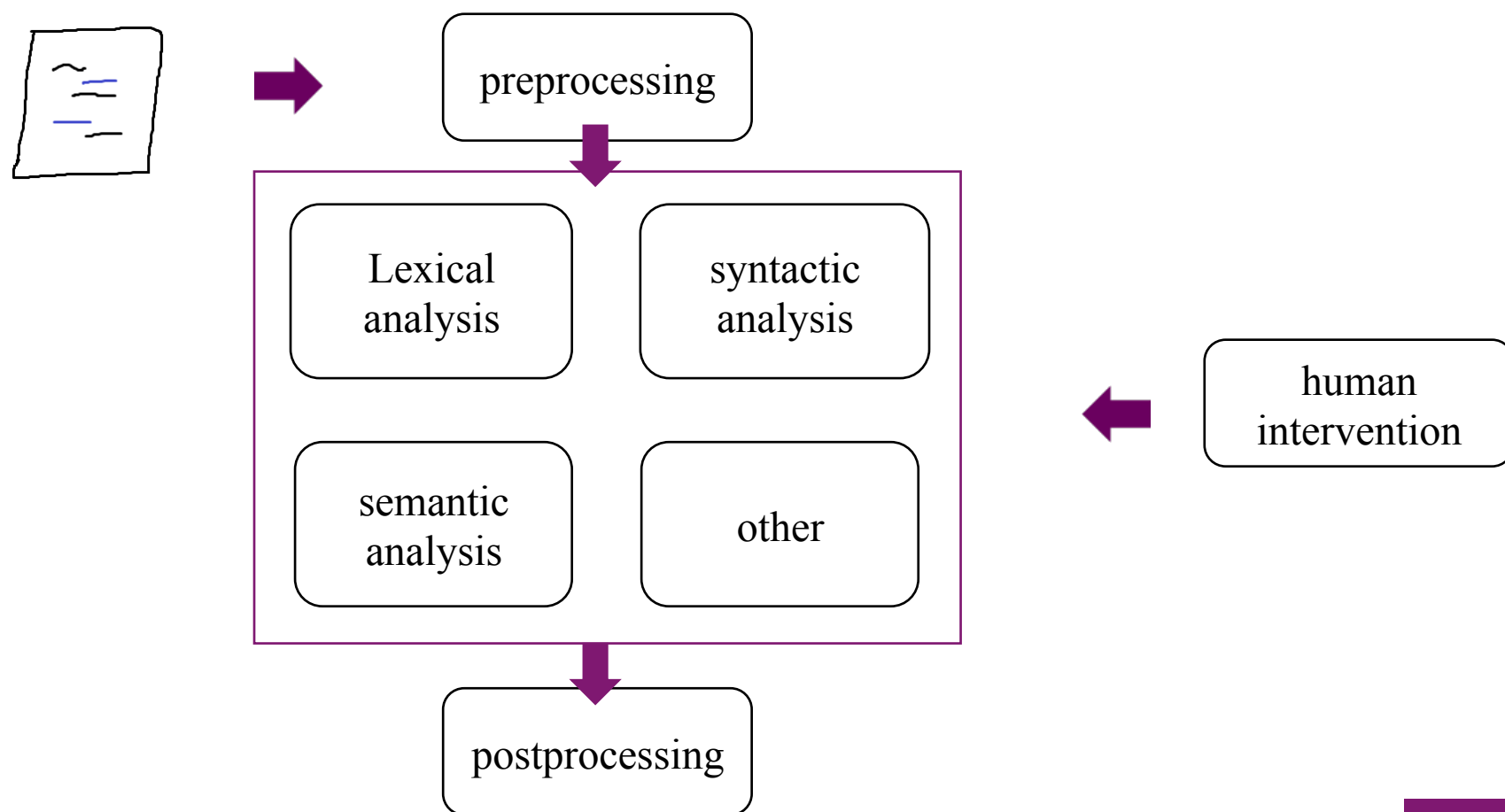> RQ4: What is the state of art of applying NLP in software engineering?

Current Work——An SLR for NLP

Senarios

> **Requirement documents**：requirements prioritization , mapping of requirements to formal representations, feature location, test cases generation, ……

> **developers' communication**：mining software-specific terms, ……

> **Bug reports**：duplication detection, similar bugs recommend, ……

> **Other**：API documents, event logging, user reviews,……

Current Work——An SLR for NLP

Process Model

CONTENTS

Future Work

Apply NLP in Grey Literature Review

- In which step can NLP be used in grey literature review ?

- What techniques and tools can be used according to different tasks ?

- Is there any potential improvement while applying NLP ?

THANKS

Jinyu Ma
mg1732011@smail.nju.edu.cn