

What is Natural Language Processing (NLP)?

<https://powcoder.com>

Assignment Project Exam Help

From Wikipedia: "Natural language processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data."

What is Computational Linguistics (CL)?

<https://powcoder.com>

From the ACL website: “Computational linguistics is the scientific study of language from a computational perspective.

Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or 'empirical'). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system.”

What is the relation between CL and NLP?

<https://powcoder.com>

- ▶ Most of the time the two terms are used interchangeably.
- ▶ In practice, some researchers are more interested to build computational models as an explanation of some linguistic phenomena ("Can the structure of a sentence be represented with a tree?"), while others are more interested in building a working system or an NLP application (a Machine Translation system (MT), a dialogue system)
- ▶ There are more and more of the latter types as these applications are put to everyday use. Have you ordered anything with Alexa?

NLP Applications

<https://powcoder.com>

Assignment Project Exam Help

- ▶ Sentiment, opinion, emotion analysis
- ▶ Information Extraction, Knowledge Acquisition
- ▶ Question Answering, Machine Reading
- ▶ Machine Translation
- ▶ Text summarization
- ▶ Spoken Language Understanding, Dialogue systems
- ▶ Many others

Sentiment Analysis

<https://powcoder.com>

Sentiment Analysis Assignment Project Exam Help

Sentiment Analysis predicts whether an input is positive or negative. [View More](#)

Enter text or
Choose an example...

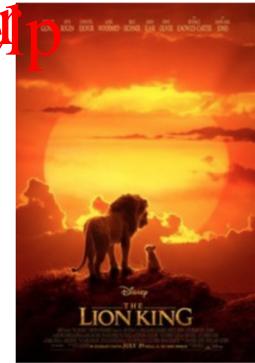
Input

It was good if you're tone deaf, I suppose. The song was littered with inappropriate lyrics, but that could have been intentional in tracking creativity and originality. The opening did appear to

Run >

Answer
Negative

<https://powcoder.com>



THE LION KING

Input

made me cry during the opening scene bringing a little bit of nostalgia to the movie with a mix of new voices and characters

Run >

Answer
Negative

[Add WeChat powcoder](https://powcoder.com)

Information Extraction

<https://powcoder.com>

Detected language: English

Show me the response

Show API url

The **Mona Lisa** is a 15th-century oil painting created by **Leonardo**. It's held at the **Louvre** in **Paris**.

1 person

1 work

10 projects

2 places

10 events

1 concept



Mona Lisa



Oil painting



Leonardo da Vinci



Louvre



Paris

Machine Translation

<https://powcoder.com>

Assignment Project Exam Help

Assignment Help Example

Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical"). Working in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of speech or natural language system.

https://powcoder.com Add WeChat powcoder

计算语言学是从计算角度对语言进行科学的研究。计算语言学家有兴趣提供各种语言现象的计算模型。这些模型可以是“基于知识”（“手工制作”）或“数据驱动”（“统计”或“经验”）。计算语言学在某些情况下是由科学观点驱动的，其中一个人试图为特定的语言或心理语言现象提供计算解释；在其他情况下，动机可能更纯粹是技术，以提供语音或自然语言系统的工作组件。

Jìsuàn yǔyán xué shí cóng jìsuàn jiāodù duì yǔyán jīnxíng kǒuxué yánjiū. Jìsuàn yǔyán xué jiā yóu qíngxù tǐgōng gè zhǔnghóng yuán xiǎngxiàng de jìsuàn móxíng. Zhexiē móxíng kěi shí "jíyú chǐshí" (jíyú gǔtǐng zhǐshí) yě néng shì "tāngtǐng" (tāngtǐng "tōngjì") huò "jīngyàn". Jìsuàn yǔyán xué zài mǎlù le jīngqíng kǎixuǎn yuán xiāngxiàng de qǐngkǎi qíngkǎi rén shí shù wéi tǐdèng de yǔyán huò xīnlǐ yǔyán xiānxiàng tǐgōng jìsuàn jiēshí; zài qítā qíngkuàng xià, dōngjī kēngēng gèng chǔncuī shí jīshù, yí tǐgōng yǔyīn huò zírán yǔyán xiāngtóng de gōngzuò zūjiān.

[Show less](#)

Machine Translation

<https://powcoder.com>

Assignment Project Exam Help

≡ Google Translate

The screenshot shows the Google Translate interface. The source text is in Chinese: "计算语言学是从计算角度对语言进行科学研究。计算语言学家有兴趣提供各种语言现象的计算模型。这些模型可以是“基于知识”（“手工制作”）或“数据驱动”（“统计”或“经验”）。计算语言学在某些情况下是由科学观点驱动的，其中一个人试图为特定的语言或心理语言现象提供计算解释；在其他情况下，动机可能更纯粹是技术，以提供语音或自然语言系统的工作组件。" The target text is in English: "Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models for various linguistic phenomena. These models can be "knowledge-based" ("handmade") or "data-driven" ("statistics" or "experience"). Computational linguistics is in some cases driven by a scientific perspective in which one attempts to provide a computational interpretation of a particular language or its linguistic phenomena; in other cases, the motive may be more purely technical to provide a speech or natural language system. Working component." The interface includes language detection, a toolbar with text, document, and image options, and a feedback button.

Machine Reading

<https://powcoder.com>

ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE
Reading Comprehension

Reading comprehension is the task of answering questions about a passage or text to show that the system understands the passage.

Assignment Project Exam Help

Enter text or
Choose an example... ▾

Passage

trying to provide a computational explanation for something that one may not have known; in some cases the motivation may be purely technical; that one wants to provide a working component of a speech or natural language system; and in the work of computational linguists, incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.

Question

what is computational linguistics?

Assignment Project Exam Help

Model
 BiDAF (trained on SQuAD)
 NAQANet (trained on DROP)

Run >

Add WeChat powcoder

Answer
scientific study of language

Passage Context

Computational linguistics is the [scientific study of language](#) from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical"). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system. Indeed, the work of computational linguists is incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.

Machine Reading

<https://powcoder.com>

 ALLEN INSTITUTE for ARTIFICIAL INTELLIGENCE

Reading Comprehension

Assignment Project Exam Help

Enter text or
Choose an example... ▾

Passage

trying to provide a computational explanation for linguistic phenomena and how it uses them to answer questions about the text. In one wants to provide a working component of a speech or natural language system and the work of computation linguistics is incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.

Question

What are the systems that the work of computational linguists is incorporated into?

Model

BiDAF (trained on SQuAD)
 NAQANet (trained on DROP)

Add WeChat powcoder Run >

Answer

working systems 

Passage Context

Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical"). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system. Indeed, the work of computational linguists is incorporated into many **working systems** today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few.

Question Answering

Google Who is the first president of Brandeis University?  

All News Images Shopping Maps More Settings Tools

Assignment Project Exam Help

About 11,600,000 results (1.19 seconds)

Brandeis University Founder
Assignment Project Exam Help

Abram L. Sachar

<https://powcoder.com>

Add WeChat powcoder

Abram Leon Sachar was an American historian and founding president of Brandeis University.
[Wikipedia](#)

People also search for

| | | | |
|--|--|---|--|
|  Howard Sachar Son |  Louis Brandeis |  Eleanor Roosevelt |  Luther Rice |
|--|--|---|--|

Text Summarization

Apple's [new titanium credit card](#) comes with a potential headache that will be familiar to iPhone users everywhere: It can get scuffed up a little too easily in the real world.

Apple ([AAPL](#)) updated its [support](#)

[webpage](#) this week to warn customers that some fabrics, such as leather and denim, could cause permanent discoloration to the Apple Card. That damage won't wash off. Apple also advises against placing the card in a wallet slot that already has a different credit card, so it doesn't get scratched. And the company says customers should not store their Apple Card in a pocket or bag with loose change, keys or "other potentially abrasive objects."

"If your titanium Apple Card comes into contact with hard surfaces or materials, it's possible that the coating can be damaged," Apple said on its website.

The bottom line: The Apple Card probably shouldn't touch anything.

<https://powcoder.com>

This is a webinterface to the [Open Text Summarizer](#) tool. The various languages and tries to identify the most important parts

Just paste your text or load it from an URL to get it summarized.

[Assignment Project Exam Help](#)

<https://powcoder.com>

[Add WeChat powcoder](#)

Input

should not store their Apple Card in a pocket or bag with loose change, keys or other potentially abrasive objects." "If your titanium Apple Card comes into contact with hard surfaces or materials, it's possible that the coating can be damaged," Apple said on its website. The bottom line: The Apple Card probably shouldn't touch anything.

(or load from URL)

Output

Summary Keywords

Summarization Ratio

5% 10% 20% 30% 40% 50% 60%

Language

en 

Submit

Apple's new titanium credit card comes with a potential headache that will be familiar to iPhone users everywhere: It can get scuffed up a little too easily in the real world.

Dialogue Systems

Try a multi-turn interaction <https://powcoder.com>

Say "Alexa", then ask "Set an alarm". When she asks what time, just say a number, like "8". She'll want to know if that's AM or PM.

Assignment Project Exam Help

You probably noticed that despite having a bit of back and forth with Alexa, you only said the wake word *once* at the start of the conversation. This is called a **multi-turn interaction** and it's a more natural method of communication because you can continue speaking without starting every phrase with "Alexa".

Assignment Project Exam Help

In your terminal window, you can scroll up until you see the UI state **Listening....** Right above that you'll see that the state of the **Audio Input Processor** (AIP) has changed from **IDLE** to **EXPECTING_SPEECH** and then **RECOGNIZING** - without you speaking the wake word!

<https://powcoder.com>

```
2017-11-15 17:53:42.132 [ 8] 0 CapabilityAgent:removingMessageIdFromMap:messageId=d530f5b4-fd12-46a7-be7d-41ebae330
2017-11-15 17:53:42.132 [ 10] 1 DirectiveRouter:handleDirective:messageId=4f22bdfc_e27-4f7d-980d-01856aa6fc19,action
2017-11-15 17:53:42.133 [ 9] 0 AudioInputProcessor:setState:from=IDLE,to=EXPECTING_SPEECH
2017-11-15 17:53:42.134 [ 91] 0 DirectiveRouter:directiveHandler:listeningCompleted:messageId=4f32bdfc_e27-4f7d-980d-01856aa6fc19
2017-11-15 17:53:42.134 [ 9] 0 CapabilityAgent:removingMessageIdFromMap:messageId=d530f5b4-fd12-46a7-be7d-41ebae330
2017-11-15 17:53:42.134 [ 9] 0 AudioInputProcessor:setState:from=EXPECTING_SPEECH,to=LISTENING
2017-11-15 17:53:42.134 [ 9] 0 AudioInputProcessor:setState:from=LISTENING,to=RECOGNIZING
#####
# Listening... #
#####
```

Alexa only listens when customers indicate that they want to speak to the cloud. Typically, this means the AIP was triggered by the **Wake Word Engine** running on the client. In this case, it's been activated via a **Directive** delivered down to your client from the cloud. When Alexa asks **you** a question, the AIP activates because it knows the customer wants to provide a response.

You can learn more about multi-turn interactions [here](#).

NLP Tasks

<https://powcoder.com>

Many of these problems are complex and cannot be solved with a single model. So they are decomposed into smaller, self-contained problems that can be solved individually and then chained together into a pipeline system. These problems are called **tasks** in NLP convention.

Take Information Extraction (IE), for example.

- ▶ <https://powcoder.com>
- ▶ Named Entity Recognition (NER)
- ▶ Coreference resolution
- ▶ Relation Extraction
- ▶ Knowledge base population

Pipeline systems are susceptible to **error propagation**. Anytime you can combine individual tasks and do **joint inference**, you can usually improve system performance.

NLP Tasks are often influenced by linguistic conceptions

<https://powcoder.com>

Assignment Project Exam Help

| Linguistic layer | NLP task |
|------------------|--|
| Morphology | Word tokenization/segmentation, morphological analysis |
| Syntax | POS tagging, Syntactic parsing (dependency/constituency) |
| Semantics | Semantic role labeling, meaning representation parsing |
| Discourse | Discourse parsing |
| Pragmatics | Dialogue act tagging |

Formal characterization of NLP tasks

<https://powcoder.com>

- ▶ Simple classification problems
 - ▶ Sentiment/opinion/emotion analysis, text classification, word sense disambiguation (WSD), etc...
- ▶ Sequence labeling problems
 - ▶ Tokenization, POS tagging, NP-chunking, NER, code switching, dialogue act tagging, Multi-word expression (MWE) detection, etc...
- ▶ Problems that can be modeled as trees and graphs
 - ▶ Syntactic (dependency and constituent) parsing, meaning representation parsing
- ▶ Sequence-to-sequence problems
 - ▶ Machine Translation, Text Summarization, dialogue systems(?)

Steps in developing an application

<https://powcoder.com>

Assignment Project Exam Help

Supposed you are asked (or wanted) to develop an NLP application, you need to think about

- ▶ How do I decompose the application into solvable tasks, given the current state of the art?
- ▶ For each task, what is the most appropriate machine learning method for each of the tasks?
- ▶ What type of training data should I create (purchase, license)?

General formulation of Learning

<https://powcoder.com>

Many NLP problems can be formulated mathematically as optimization:

Assignment Project Exam Help

$$\hat{y} = \operatorname{argmax}_{y \in \mathcal{Y}} \Psi(x, y; \theta)$$

Assignment Project Exam Help

where,

<https://powcoder.com>

- ▶ x is an element of a set \mathcal{X}
- ▶ y is an element of the set $\mathcal{Y}(x)$
- ▶ Ψ is a scoring function or a **model**, which maps from the set $\mathcal{X} \times \mathcal{Y}$ to real numbers
- ▶ θ is a vector of parameters for Ψ
- ▶ \hat{y} is the predicted output, which is chosen to maximize the scoring function

Learning and search

<https://powcoder.com>

- ▶ Search is the procedure of finding the output \hat{y} that gets the best score with respect to the input x by computing the argmax of the scoring function Ψ
 - ▶ The search can be simple if it's a matter of finding the best label among a small set of labels (e.g., sentiment analysis)
 - ▶ Or it needs a non-trivial search algorithm (e.g., finding the best part-of-speech sequence for a sentence)
- ▶ Learning is the process of finding the parameters θ .
 - ▶ This is done by optimizing some function of the model and the labeled data in a training process
 - ▶ The parameters are usually continuous, and learning algorithms generally rely on **numerical optimization** to identify vectors of real-valued numbers

Learning Methods for Simple Classification

<https://powcoder.com>

- ▶ Supervised Learning
 - ▶ Generative
 - ▶ Naive Bayes
 - ▶ Discriminative
 - ▶ linear models: Logistic Regression, Perceptron, Support Vector Machines*
 - ▶ Non-linear models (neural network models or deep learning, multiple layers: MLP, CNN, RNN)
- ▶ Unsupervised Learning
 - ▶ EM-based algorithms (backward-forward, inside-outside), clustering algorithms (K-means, EM, Hierarchical)
- ▶ Semi-supervised methods
- ▶ Search is usually trivial, and involves finding the label that gets the highest score.

Sequence Labeling methods

<https://powcoder.com>

Sequence labeling methods can be viewed as classification combined with a search algorithm

Assignment Project Exam Help

- ▶ Supervised Learning
 - ▶ Generative
 - ▶ Hidden Markov Models (HMM). Naive Bayes combined with the Viterbi Algorithm
 - ▶ Discriminative
 - ▶ linear models: Conditional Random Fields (CRF) Logistic Regression combined with the Viterbi Algorithm, Perceptron combined with the Viterbi Algorithm
 - ▶ Non-linear models: LSTM-CRF, a form of RNN combined with a search algorithm
- ▶ Unsupervised Learning
 - ▶ Backward-forward, a form of EM algorithm for sequences
- ▶ semi-supervised methods

Tree-based learning algorithms

<https://powcoder.com>

Tree-based methods can also be viewed as classification combined with a search algorithm

Assignment Project Exam Help

► Supervised Learning

- Generative parsing models

► Native PBNs incorporated with the CKY algorithm for constituent parsing

► Discriminative

- Perceptron combined with Beam Search for dependency parsing
- Perceptron or logistic regression combined with greedy search for constituent parsing

► Non-linear models:

- LSTM combined with CKY

► Unsupervised Learning

- Inside-outside, a form of EM algorithm for trees

► semi-supervised methods

Learning and linguistic knowledge

- ▶ The relative importance of learning and linguistic knowledge has been a recurring topic of debate.
 - ▶ “Every time I fire a linguist, the performance of our speech recognition system goes up.”
- ▶ Linguistic knowledge figures prominently in early rule-based systems
- ▶ Statistical systems in the 1990s and early 2000s focus on linguistically inspired tasks with carefully engineered features
- ▶ Deep learning methods enabled end-to-end systems
 - ▶ Particularly effective in areas like MT where there are large scale training sets
 - ▶ “Natural language processing from scratch”
 - ▶ Model architectures are still inspired by linguistic theories
- ▶ The debate is far from being settled. Linguistic knowledge is particularly needed in problems that require “deep understanding”.

What type of math is needed

<https://powcoder.com>

- ▶ Probabilities
 - ▶ The output of a classifier is often expressed in terms of a probabilistic distribution: “This review has a 82% probability of being positive, and a 18% probability of being negative”
 - ▶ For some models (e.g., Naive Bayes), the parameters Θ are in the form of probabilities
- ▶ Calculus: know how to compute derivatives for various functions
 - ▶ The most basic form of machine learning is optimization based on gradient descent/ascent to achieve the maximums or minimums
- ▶ Linear algebra: increasingly, you have to manipulate vectors and matrices (or “tensors”)
- ▶ Some of you may not have calculus. We plan to have some tutorials on this topic.

Where do you need linguistics?

<https://powcoder.com>

- ▶ Most of the NLP tasks rely on linguistic concepts that are intuitive.
Assignment Project Exam Help
 - ▶ The part speech of a word: verbs, nouns, adjectives
 - ▶ Named entities: person, organization, geographic entities, geopolitical entities, etc.
 - ▶ Reviews: This is a positive review
- ▶ Some linguistic concepts require some formal linguistic training
Assignment Project Exam Help
 - ▶ Syntactic trees (constituent or dependent)
- ▶ When breaking down an application into smaller tasks that you can solve with machine learning, you need a good understanding of different layers of analysis: morphology, syntax, semantics, discourse and dialogues, etc.
Add WeChat powcoder
- ▶ You also need good linguistic intuition to come up with good features or architectures in your statistical model

What kind of programming skills do you need?

<https://powcoder.com>

Assignment Project Exam Help

- ▶ Python, Python, Python!
- ▶ Python libraries:
 - ▶ numpy, Tensorflow (version 1.12 or 1.13), Keras (now part of Tensorflow), PyTorch, MXNet
 - ▶ We are going to use PyTorch for some of our projects
- ▶ We are holding tutorials on PyTorch for people who are new to this during recitation.

Course requirements

<https://powcoder.com>

- ▶ Prerequisites
 - ▶ CS114 or permission by instructor (talk to me if you're not sure you should take the course)
 - ▶ Programming experience: familiar with Python and Python packages such as Numpy and PyTorch
 - ▶ Statistics/machine learning/math background and ability to pick up some linguistics
- ▶ Disciplines Relevant to NLP
 - ▶ Linguistics
 - ▶ Computer science, Artificial Intelligence
 - ▶ Machine learning

Textbooks

<https://powcoder.com>

- ▶ Textbooks
 - ▶ Required: Introduction to Natural Language Processing by Jacob Eisenstein (pre-publication version can be downloaded from <https://github.com/jacobeisenstein/gmp-class/blob/master/notes/eisenstein-nlp-notes.pdf>)
 - ▶ Recommended: Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola: Dive into Deep Learning (Online book, can be accessed from <https://d2l.ai>)
- ▶ Supplemental online material
 - ▶ PyTorch tutorials:
https://pytorch.org/tutorials/beginner/deep_learning_60min_blitz.html
 - ▶ PyTorch Documentation:
<https://pytorch.org/docs/stable/index.html>

Perusall

<https://powcoder.com>

Assignment Project Exam Help

- ▶ I plan to experiment with Perusall: <https://app.perusall.com/>
- ▶ Perusall allows users to highlight specific areas of a PDF document to enter your comments
- ▶ When possible, I plan to post the slides to Perusall before each the class so that you can enter comments and questions. I can then address those questions during the class.
- ▶ To access the course, enter the access code: XUE-XPWJB

Course work

<https://powcoder.com>

- ▶ Participation: 5%
 - ▶ Credits for active participation in class discussions, and for asking questions.
- ▶ 4 Projects (50%)
 - ▶ Projects are more open-ended, though starter code is provided for some projects. Projects require experiments and write-up.
- ▶ 3-4 Assignments (30%)
 - ▶ Programming assignments for well-defined problems
- ▶ 1 final Quiz (15%)
 - ▶ The quiz tests important concepts covered in the course.
- ▶ **Academic integrity:** You should finish homework assignments, exams, and project reports on your own unless a project is explicitly stated as a collaborative project. Late projects are subject to grade deduction.