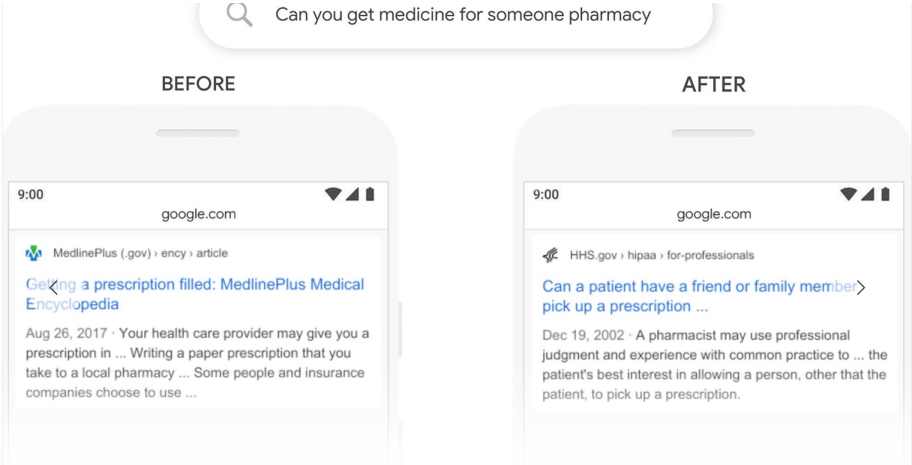
Natural Language Processing

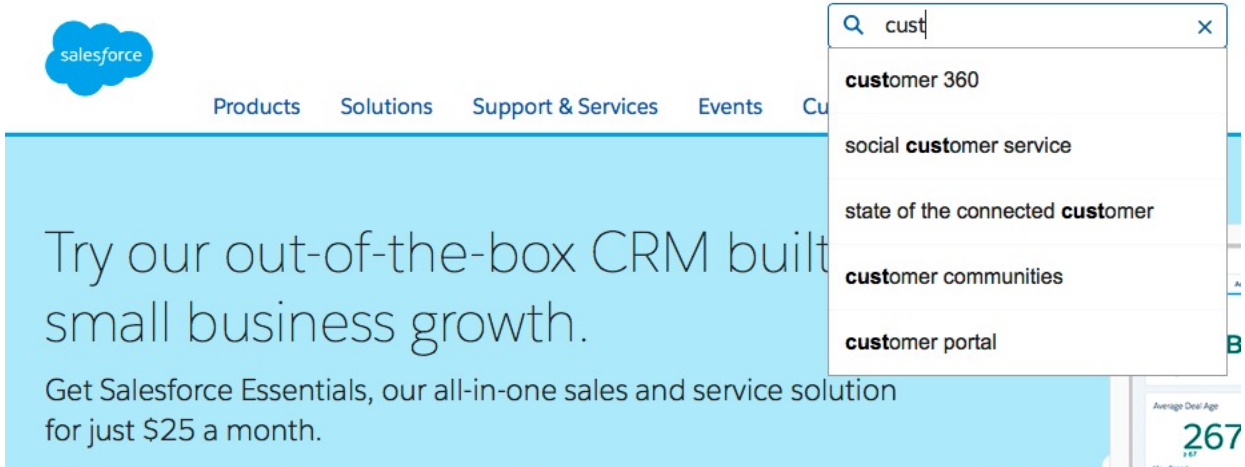
Unit 4

Working in NLP

Faizan Javed
DataScience @ SMU

Industry use cases – Web and Enterprise Search





Industry use cases – Healthcare

Using Amazon Comprehend Medical in the AWS Console

Opening the AWS Console, all we have to do is paste some text and click on the 'Analyze' button.

Input text

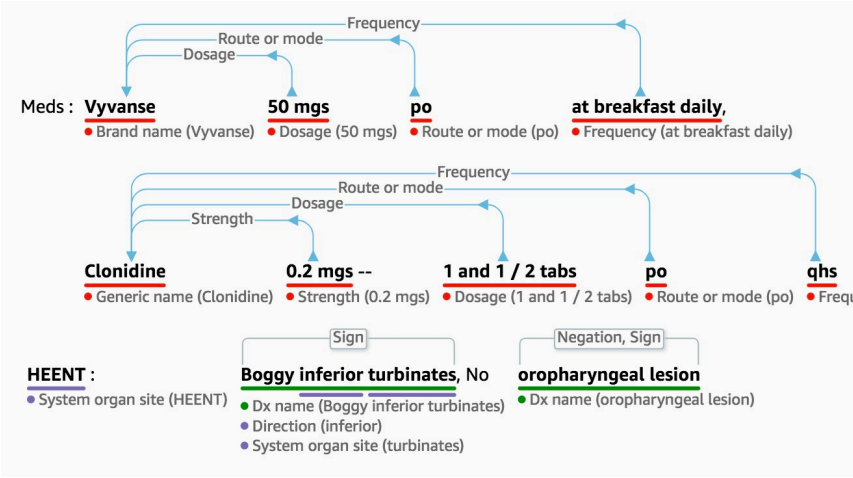
Pt is 40yo mother, software engineer
HPI : Sleeping trouble on present dosage of Clonidine. Severe Rash on face and leg, slightly itchy
Meds : Vyvanse 50 mgs po at breakfast daily, Clonidine 0.2 mgs -- 1 and 1 / 2 tabs po qhs
HEENT : Boggy inferior turbinates, No oropharyngeal lesion. Lungs : clear. Heart : Regular rhythm. Skin : Papular mild erythematous eruption to hairline
Follow-up as scheduled

407 of 10000 characters used.

Clear text | Analyze

The document is processed immediately. Entities are extracted and highlighted: we see personal information in orange, medication in red, anatomy in purple and medical conditions in green.

40yo 0.99+ score	Age	Protected health information	-
software engineer 0.98 score	Profession	Protected health information	-
Sleeping trouble 0.81 score	Dx name	Medical condition	Symptom 0.68 score
Clonidine 0.98 score	Generic name	Medication	-
Rash 0.99+ score	Dx name	Medical condition	Symptom 0.75 score



Industry use cases – Legal

Lexis Advance®

What is the doctrine of last c

Suggested Questions

- What is the doctrine of last clear chance?
- What is the doctrine of court of last resort?
- What is the doctrine of last commanded?

Content Type | Federal | State | Practice Area or Industry | International

Cases | Secondary Materials | News

Statutes and Legislation | Forms | Legal News

Select Category
Cases 5,689

Results for: fraud on the market | Actions

Filters | Clear

2. Amgen Inc. v. Conn. Ret. Plans & Trust Funds | Supreme Court of the United States | 2013-02-27 | 568 U.S. 455

Sort by: Court

SUPREME
CIRCUIT
DISTRICT
STATE

1985 1990 1995 2000 2005 2010 2015

2. Amgen Inc. v. Conn. Ret. Plans & Trust Funds

Feb 27, 2013 | 568

Distinguished by
54. Halliburton Co. v. Erica P. John Fund, Inc. | Supreme Court of the United States | 2014-06-23 | 134 S. Ct. 2398

Although section 10(b) does not create an express private cause of action, we have long recognized an implied private cause of action to enforce the provision and its implementing regulation. See Blue Chip Stamps v. Manor Drug Stores, 421 U.S. 723, 730, 95 S. Ct. 1917, 44 L. Ed. 2d 539 (1975). ... To recover damages for violations of section 10(b) and Rule 10b-5, a plaintiff must prove "(1) a material misrepresentation or omission by the defendant; (2) scienter; (3) a connection between the misrepresentation or omission and the purchase or sale of a security; (4) reliance upon the misrepresentation or omission; (5) economic loss; and (6) loss causation." Amgen Inc. v. Connecticut Retirement Plans and Trust Funds, 568 U.S. ___, ___, 133 S. Ct. 1184, 185 L. Ed. 2d 308, 316 (2013) (quoting Matrixx Initiatives, Inc. v. Siracusano, 563 U.S. ___, ___, 131 S. Ct. 1309, 179 L. Ed. 2d 398, 409 (2011)).

guisite under
of a class of
action, since
evidence
of proof of
rather than

9, 1993 | 5 Cal. 4th 1

ock and
epresentations
could not bring an action for deceit under state securities law without alleging actual reliance on defendants' misrepresentations.

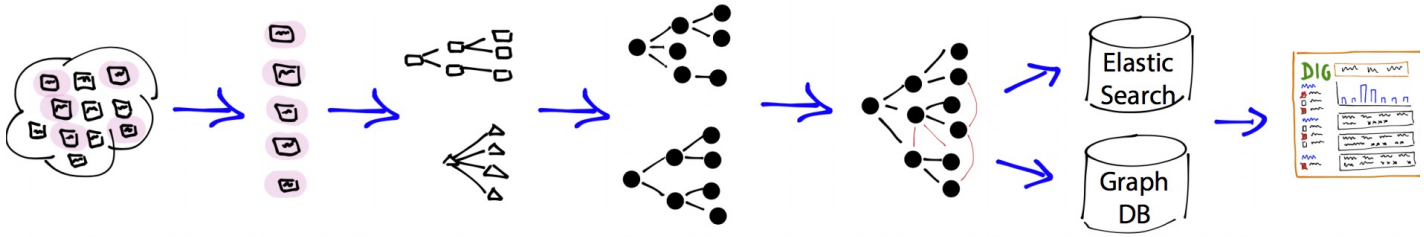
4. In re HealthSouth Corp. Sec. Litig.

United States District Court for the Northern District of Alabama, Southern Division | Mar 31, 2009 | 257 F.R.D. 260

Overview: In securities fraud action, stockholders established by a preponderance of evidence Fed. R. Civ. P. 23(a) and (b)(3) requirements for class certification. Court found no reason why Basic fraud-on-market presumption of reliance should not apply here, and investment bankers did not rebut that

Industry use cases – Law enforcement/human trafficking

REUSABLE TECHNOLOGY FOR BUILDING DOMAIN-SPECIFIC SEARCH



CRAWLING | EXTRACTION | MAPPING TO ONTOLOGY | ENTITY LINKING | KNOWLEDGE GRAPH DEPLOYMENT | QUERY & VISUALIZATION

Industry use cases – Law enforcement/human trafficking

DIG Home Saved Queries Notifications Hello memex

Filter Folders

FROM

TO

CITY/REGION

Not Specified	663544
Toronto	111130
New York	81876
Ontario	80422
Los Angeles	54444
Chicago	54285
Phoenix	49934
Montreal	48932
Denver	48291
Houston	47814
Calgary	47480
Red Deer	46936
San Jose	46769
Ottawa	46367
Dallas	45777
London	43122
Orange County	43001
San Diego	42382
Boston	42352
Edmonton	42097
Atlanta	41958
Display a menu	41155

jessica red hair blue eyes swedish | Search | Clear All | Save

With Images Only

SORT BY Best Match

0 item(s) selected

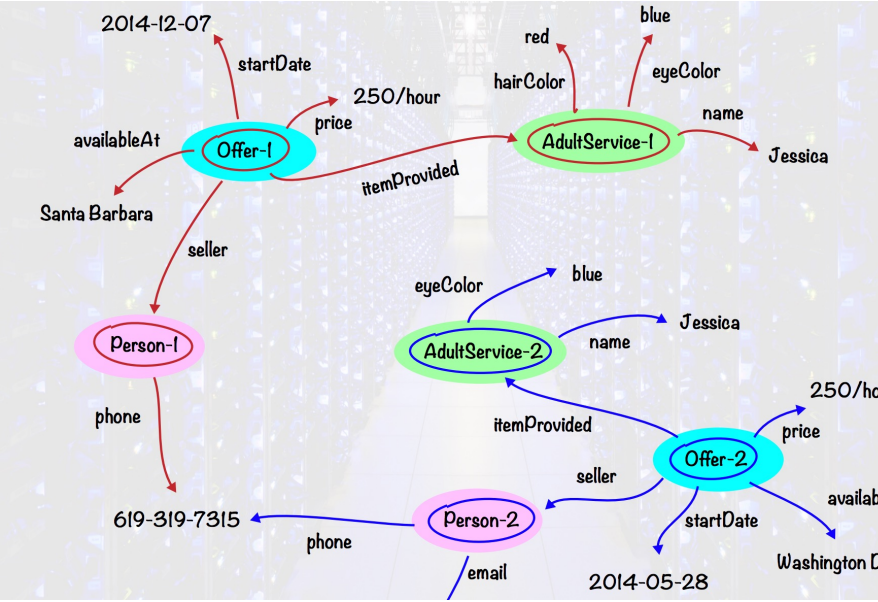
4,946,783 results

San Diego Escorts - Elite Swedish Escort in San Diego 31 - 3518804 - My Provider Guide
created: 04/11/2011 00:00:00 UTC location: San Diego phone: +1-6193197315
name: Jessica age: 31

1-619-319-7315 - jessica @swedishgfe.com - swedishgfe - San Diego Escorts - Swedish Pleasure from a Real Swede - 3558086 - My Provider Guide
created: 01/14/2014 00:00:00 UTC location: San Diego phone: +1-6193197315
name: Jessica age: 33

1-619-319-7315 - jessica @swedishgfe.com - swedishgfe - San Diego Escorts - Swedish Pleasure from a Real Swede - 3558086 - My Provider Guide
created: 01/14/2014 00:00:00 UTC location: San Diego phone: +1-6193197315
name: Jessica age: 33

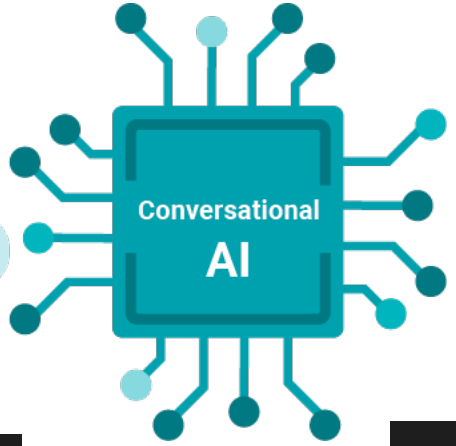
1-619-319-7315 - jessica @swedishgfe.com - swedishgfe - San Diego Escorts - Swedish Pleasure from a Real Swede - 3558086 - My Provider Guide
created: 11/01/2013 00:00:00 UTC location: San Diego phone: +1-6193197315
name: Jessica age: 33



Industry use cases – Conversational AI

ref. kore.ai

- Can you help me (*service request*)
- I have a question (*Information request*)
- I need to get this done (*action request*)
- I want to see this report (*info request*)
- What should I do (*decision request*)
- Let me know when X occurs (*notification request*)



- What does the person want?
- What language are they using?
- Are people, places and details mentioned?
- What is the kind of request?
- Does the user seem happy or upset?
- Have we interacted before?

Industry use cases – Conversational AI (ecommerce)

Verizon 10:03 AM 82%
Back H&M
Set Profile Picture

a minute ago

You added H&M by scanning a Kik Code

Hi 👋! Welcome to H&M on Kik

Let's get to know your style with a few quick questions! 💁

Do you want to see men's or women's clothing?

Men's

Great, lets get started!! 👕👞

Which of the following best describes you?

30s

Verizon 10:04 AM 82%
Back H&M
Set Profile Picture

Classic

Perf! I've just created a custom style profile for you

Here's how I work:
1. Tell me an item 👕
2. We'll send you an outfit inspo 👕👖👟
3. Shop or share it!

Choose an item to start with:

Tap a message

Sweatshirt	Sneakers
T-shirt	Polo shirt
Jeans	Chinos
Hooded jacket	Backpack
High tops	Pants

Verizon 10:04 AM 82%
Back H&M
Set Profile Picture

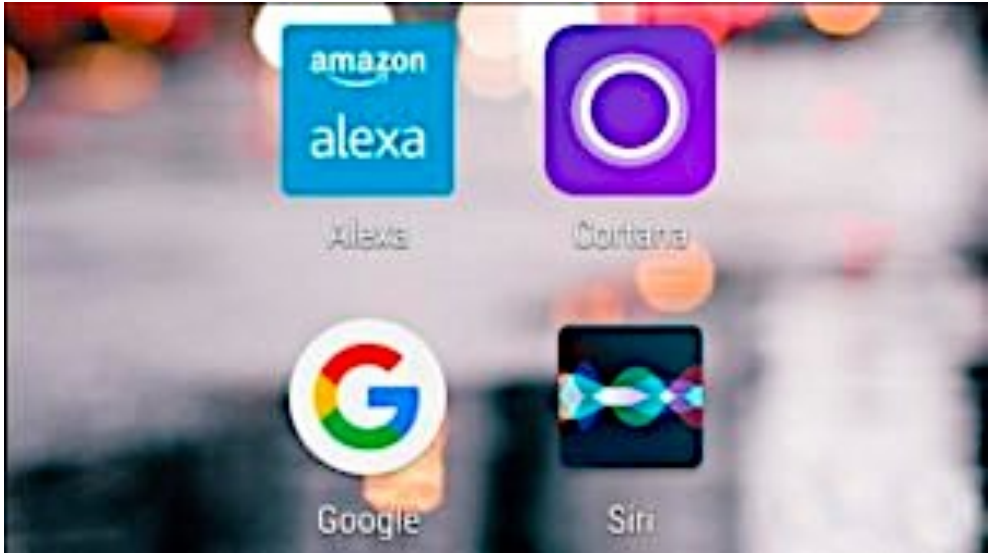
2. We'll send you an outfit inspo 👕👖👟
3. Shop or share it!

Choose an item to start with:

T-shirt

Here's an outfit with a t-shirt. Tell me what you think?

Virtual Assistants



Popular Voice Assistants and Features

Smartphone-Based Voice Assistants

Apple Siri

First mainstream voice assistant

Default wake word
Hey Siri

Common tasks
Send messages, set reminders, create calendar entries, dial calls, answer questions, set timers, show weather forecast, take photos, set alarms, play music, send texts, open apps

Google Assistant

Debuted May 2012 as Google Now in Android 4.1 aka Jelly Bean, Google Assistant debuted in 2016

Default wake word
OK Google

Common tasks
Send messages, set reminders, create calendar entries, dial calls, answer questions, set timers, show weather forecast, open apps, set alarms, play music, send texts

Smart Speaker-Based Voice Assistants

Amazon Alexa

Debuted in November 2014 in Amazon Echo and Amazon Echo Dot smart speakers

Default wake word
Alexa

Common tasks
Makes calls, music playback, sets alarms, answers questions, provides real-time information, controls smart devices using home automation systems, uses Skills apps to do more advanced features

Google Home

Debuted in November 2016

Default wake word
Hey Google

Common tasks
Answers questions, plays music, controls smart devices using home automation, makes calls, sets reminders, reads news, works with Chromecast to stream entertainment

Apple HomePod

Debuted February 2018

Default wake word
Hey Siri

Common tasks
Plays music, controls devices using smart home automation, sets timers, reads news and information, creates lists, sends messages, makes calls on an iPhone.

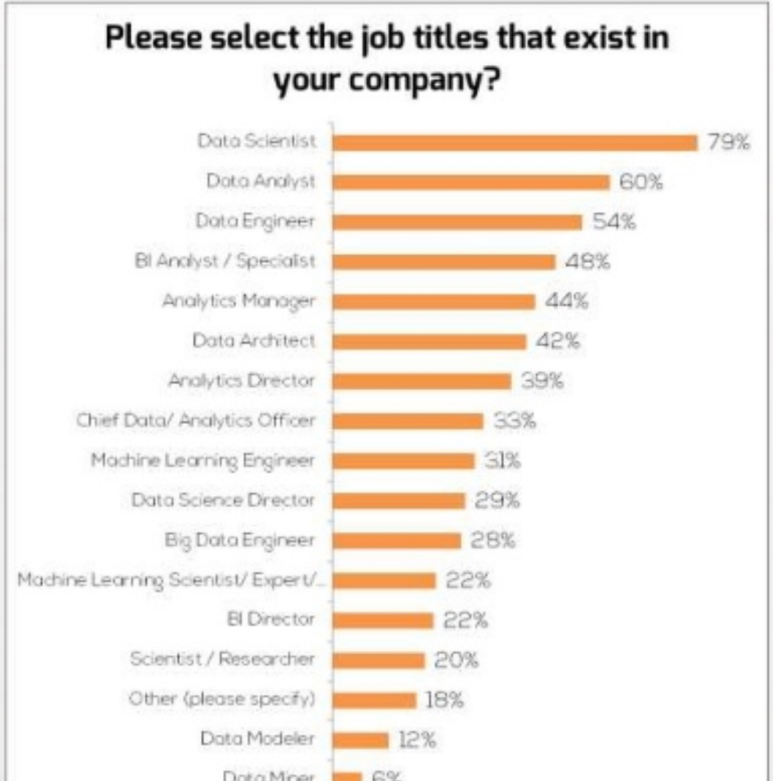
Industry use cases – Google Meena

Google Meena: Towards a Conversational Agent that Can Chat About…Anything
https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html

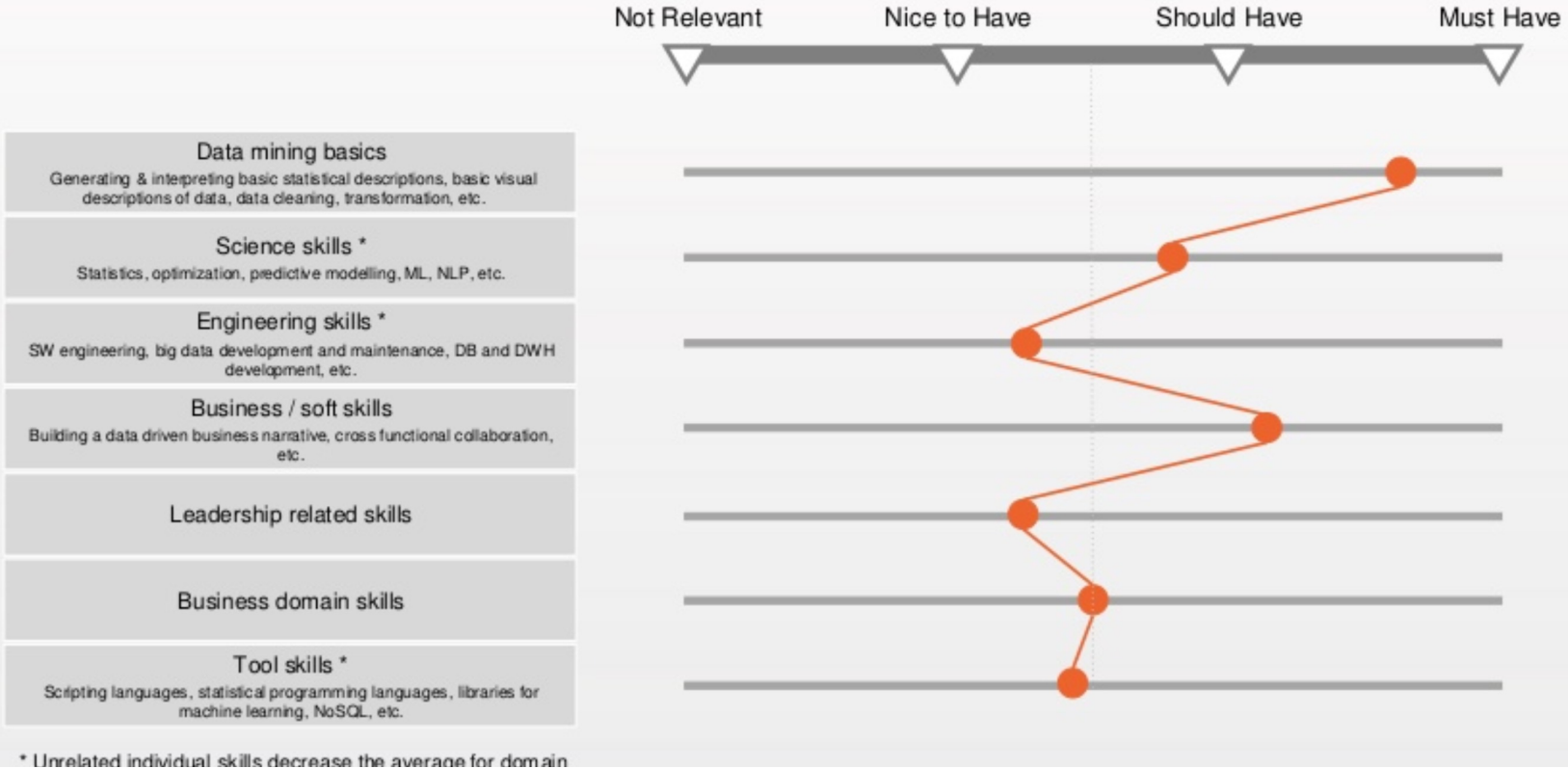
Conversation 4	
Human:	*I like how cute cows are*
Meena:	*Ah, interesting. You like how they are cute, right?*
Human:	*yeah, and they're realy smart*
Meena:	*I heard they go to college*
Human:	*cows go to college?*
Meena:	*I heard that a cow went to Harvard.*
Human:	*what did the cow study?*
Meena:	*Bovine sciences.*
Human:	*do horses go to Harvard?*
Meena:	*Horses go to Hayvard.*
Human:	*that's a pretty good joke, I feel like you led me into it*
Meena:	*You were trying to steer it elsewhere, I can see it.*

Jobs using NLP

Organizations have a multitude of titles in data science and analytics teams



Initial insights for “Data Scientist” role



Initial insights for "Data Scientist" role

Some domains in detail

Science Skills

Statistics	3.58
Optimization	2.74
Predictive modeling	3.53
Machine learning	3.16
NLP / text mining	2.19
Other domain-specific quant. skills	2.30
Top level research background	2.07

Engineering Skills

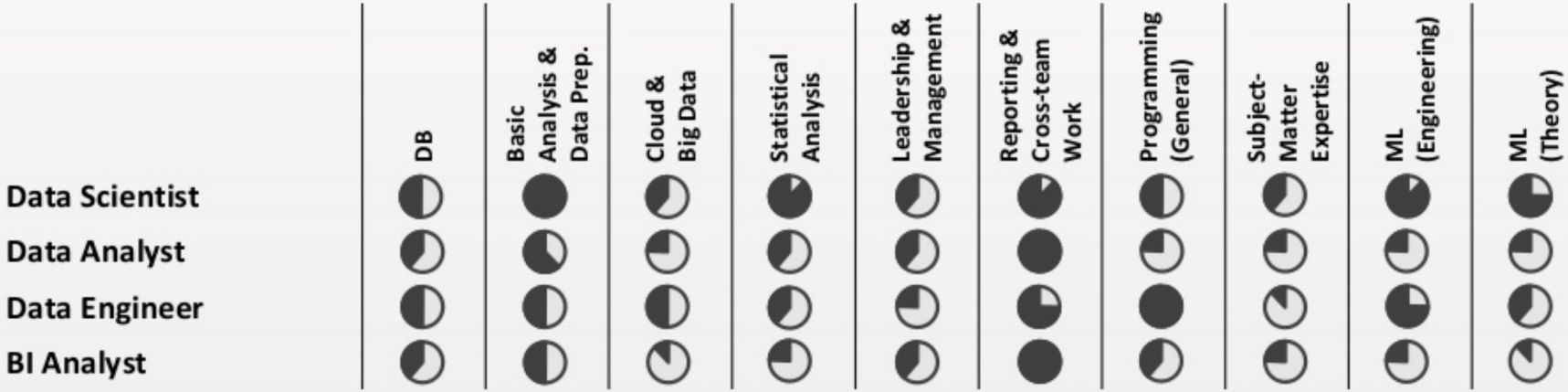
Software engineering skills	2.42
Big data development and maintenance	186
Database and DWH development and maintenance	193
Computer science fundamentals	2.65
Development and maintenance of data stores	172

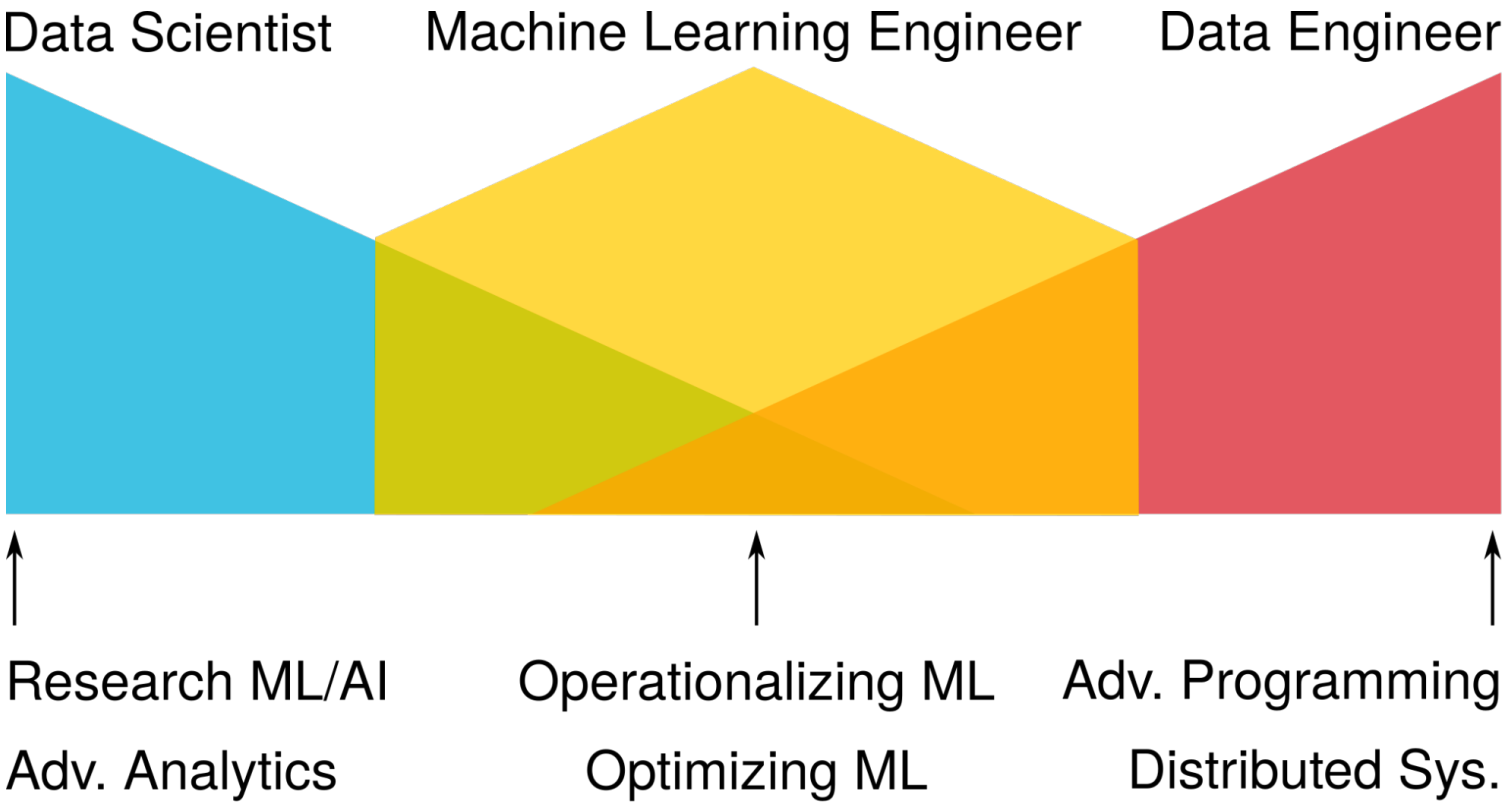
Tool Families

Programming	2.86
Database	2.35
Big Data	184
BI & Advanced Analytics	2.53
Cloud Computing	198
General Computing	193

1 = Not Relevant
2 = Nice to Have
3 = Should Have
4 = Must Have

Average role profiles as compositions of prototypical skill-sets





BDI

BIG DATA INSTITUTE

For more information go to http://bigdatainstitute.io

Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) Licensed

Data Scientist - NLP

Job ID: 839663 | Amazon.com Services, Inc.

DESCRIPTION

There's a revolution brewing to bring convenience, simplicity, efficiency, and fun to our lives, and you can be a part of it! We are hiring a Speech Scientist to be a part of the team that's creating customer experiences to discover, personalize, and acquire Alexa Skills for devices such as Amazon Echo and Amazon Fire TV. If you're not familiar with Alexa and Skills you can learn more here: https://amazon.com/alexaskills.

The Amazon Alexa Skills Store is aggressively seeking a Research Scientist with a background in NLP and wants to grow their career in that direction. As a member of a cross-functional team, you will use massive amounts of real-time and historical user data to research, develop and automate models for semantic understanding, text similarities and grammars. Working with a team of software development engineers, a successful candidate has a bias for action and is most comfortable when testing their hypotheses using heterogeneous datasets. You will be expected to challenge current methods while enriching the world's most advanced speech platform.

BASIC QUALIFICATIONS

· Masters or equivalent in a quantitative discipline of science, like computational linguistists, applied statistics, engineering, physics, computer science, economics or related field.
· 2+ years of experience in data mining and analytics techniques
· 2+ years of experience in programming in R, Python, SAS or similar languages
· Experience in NLP
· Experience coordinating with multiple teams
· Experience to performing code reviews

PREFERRED QUALIFICATIONS

· PhD or equivalent in a quantitative discipline of science, like computational linguistists, applied statistics, engineering, physics, computer science, or related field.
· Research background in NLP/NLU.
· 1+ year of experience in two or more of these or similar Natural Language Processing topics: word categorization and tagging, syntactic parsing, word sense disambiguation, topic modeling; contextual text mining, application of machine learning to NLP, semantic similarity, phrasal semantic analysis, text matching and similarity, word embedding, lexicon normalization, named entity recognition
· 1+ year of experience with NLP libraries such as NLTK, openNLP, Stanford-NLP, WordNet, SAS Text Miner or other NLP software.
· Experience working with and influencing key clients and stakeholders (internal or external).
· Experienced showing bias for action; working in a fast-paced, high-tech environment and comfortable navigating conflicting priorities and ambiguous problems.
· Experience converting ambiguous messy data into tangible actionable results.

Language Engineer

Job ID: 782833 | Amazon.com Services, Inc.

DESCRIPTION

Amazon is looking for a Language Engineer to join our Natural Language Processing (NLP) team. We are seeking a candidate with strong analytical skills and language technology experience to help us develop language components for a variety of Amazon products.

As a Language Engineer at Amazon, you will:
- Create sample language data for the target language (Arabic)
- Develop NLP components: write rules, build grammars/FSTs, create annotation systems or mappings for the target language
- Contribute to the design of new features by extending existing annotation schemas to cover new areas in the target language
- Use modeling tools to bootstrap and test new functionalities
- Analyze system performance and identify areas for improvement
- Collaborate with other Language Engineers and Data Associate teams in data collection and development efforts

BASIC QUALIFICATIONS

- Bachelor's or higher degree in a relevant field (linguistics or language)
- 2+ years experience in computational linguistics, language data processing, semantics, or syntax
- Experience with language annotation and other forms of data markup
- Experience working with speech and text language data in multiple languages
- Native or near-native fluency in Arabic

PREFERRED QUALIFICATIONS

- Experience with database queries and data analysis processes (SQL, R, Matlab, etc.), and Unix
- Experience programming in Perl, Python, or another scripting language

Applied Scientist, Natural Language Processing

Job ID: 1026105 | Amazon.com Services LLC

DESCRIPTION

Amazon is looking for a passionate, talented, and inventive Senior Scientist with a strong machine learning background to help build industry-leading Natural Language Understanding (NLU) solutions. Our mission is to push the envelope in NLU and Entity Resolution (ER) technologies to provide the best possible experience for our customers. Products such as Amazon Echo, Echo Show, and Fire TV are illustrative of the user-delighting spoken language solutions Amazon is building.

As a Senior Applied Scientist on our team, you will collaborate with other scientists and software engineers to research and develop novel algorithms and modeling techniques to advance the state of the art in NLU and ER. Your work will directly impact our customers in the form of products and services that make use of speech and language technology. You will leverage Amazon's heterogeneous data sources and large-scale computing resources to accelerate advances in natural language understanding.

Position Responsibilities:
· Participate in the design, development, evaluation, deployment of data-driven machine-learned (ML) models targeting NLU and ER applications.
· Develop and apply statistical modeling techniques (e.g. Bayesian models and deep neural networks), optimization methods, and other ML techniques to different applications in business and engineering.
· Routinely build and deploy ML models on available data.
· Research and implement novel ML and statistical approaches to add value to the business. Mentor engineers and scientists.

BASIC QUALIFICATIONS

· Master's degree in a relevant field
· 3+ years of industry experience in a relevant field or equivalent
· 5+ years of experience with data analysis and visualization tools
· 5+ years of experience in Python, or another scripting language; command line usage, e.g., Bash
· 5+ years of experience in algorithms and data structures
· 5+ years of experience with various machine learning and modeling techniques and parameters that affect their performance
· 3+ publications in scientific journals on relevant machine learning topics

PREFERRED QUALIFICATIONS

· PhD in a relevant field
· Experience in building NLU and/or ER systems (e.g. commercial speech products or government speech projects)
· Knowledge in machine learning techniques in entity resolution or text search domain.

Applied Research Scientist, Natural Language Processing

MENLO PARK, CA

APPLY TO JOB

← BACK TO JOBS

Like Share

Facebook is seeking an Applied Research Scientist to join our Language and Translation Technologies Team in Menlo Park. This team is part of the Applied Machine Learning organization within Facebook, working to unlock the meaning of language across all languages used on Facebook. In pursuit of this mission, the team has been working on creating Natural Language Processing (NLP) platforms that are used to build and deliver AI applications to more than 2 billion people that use Facebook every month. This journey is only 1% finished, and we are looking for candidates interested in extending our capabilities on cross-lingual NLP and enabling novel NLP technologies. The ideal candidate will have a strong background in developing NLP systems, a strong knowledge of ML and neural networks, and have experience working with massive amounts of data. They should also have strong software engineering skills and the ability to build systems that reach Facebook scale.

APPLIED RESEARCH SCIENTIST, NATURAL LANGUAGE PROCESSING RESPONSIBILITIES

- Conduct research to advance the state of the art in neural networks and NLP.
- Utilize that research to develop and deploy scalable neural network models into production to impact billions of people using Facebook.
- Collaborate with team members from other research and applied research teams working on a variety of deep learning and NLP problems.

MINIMUM QUALIFICATIONS

- PhD/MS with relevant experience in the fields of Natural Language Processing, Machine Learning or Deep Learning.
- 3+ years of experience in building large-scale Natural Language Processing systems from researching a prototype to production.

AI Research Engineer – Natural Language Processing

at Samsung Research America

Mountain View, CA

Title: AI Research Engineer – Natural Language Processing

Company: Samsung Research America

Lab: Samsung Artificial Intelligence Center

Location: Mountain View, CA

Lab Summary:

Samsung is working towards a vision of providing world-class AI services on its industry-leading hardware platforms like the Galaxy phones, and home appliances like the smart refrigerators and TVs, smart washers and dryers, etc. Real-time AI services on these connected devices will enable users to interact with devices more naturally via spoken dialog, visual interaction, continuous learning, and big data analysis.

Position Summary:

We are looking for candidates interested in extending our capabilities on cross-lingual NLP and enabling novel NLP technologies. The ideal candidate will have strong software engineering skills. They should also have background in NLP systems, experience of deep learning, and experience working with massive amounts of data.

Responsibilities Include:

Develop prototypes and real-world applications and improve the performance with various technologies Integrate the resulting system/software into existing platform with high quality implementation

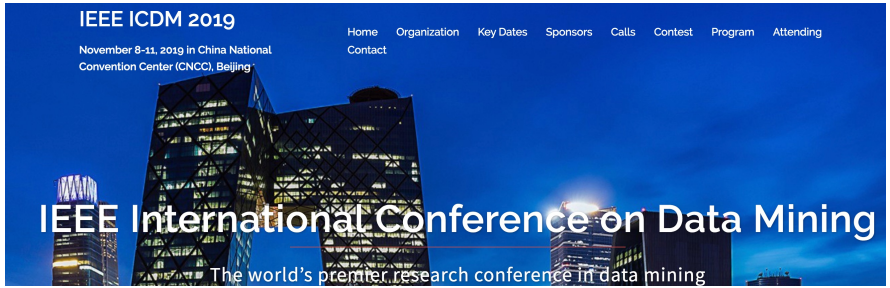
Experience Requirements:

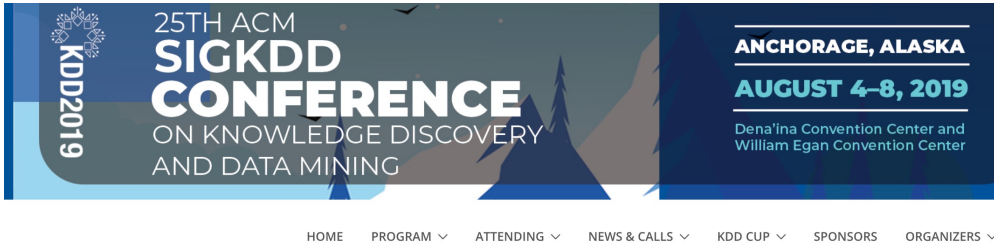
- Master or Ph.D. in Computer Science or related field
- R&D experience in natural language processing, machine learning and related fields
- R&D experience in information extraction and text mining technologies
- Hands-on experience of deep learning technologies and familiar with state-of-the-art deep learning toolkits (e.g., TensorFlow, PyTorch, Keras, etc)
- Experience in programming languages (C++, Python, Java, etc)
- Strong problem solving skills
- Good communication and team-working skills

Experience Preferred:

- 6+ years of natural language processing, information extraction, text mining technologies and related fields
- Experience in a front-end language like Elm, JavaScript, Objective-C or Swift
- Experience in at least one front-end web development framework such as React, Angular, Rails, Django, Vue.js

Conferences and organizations for NLP





CIKM 2020
19-23 OCTOBER
GALWAY · IRELAND

29TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT

Conferences and organizations for NLP

Included Themes

Methods and Algorithms

- Anomaly & Outlier Detection
- Big Data & Large-Scale Systems
- Classification & Semi-Supervised Learning
- Clustering & Unsupervised Learning
- Data Cleaning & Integration
- Deep Learning & Representation Learning
- Frequent Pattern Mining
- Feature Extraction, Selection and Dimensionality Reduction
- Mining Data Streams
- Mining Graphs & Complex Data
- Mining on Emerging Architectures & Data Clouds
- Mining Semi Structured Data
- Mining Spatial & Temporal Data
- Mining Text, Web & Social Media
- Online Algorithms
- Optimization Methods
- Parallel and Distributed Methods
- Probabilistic & Statistical Methods
- Scalable & High-Performance Mining
- Other Novel Methods

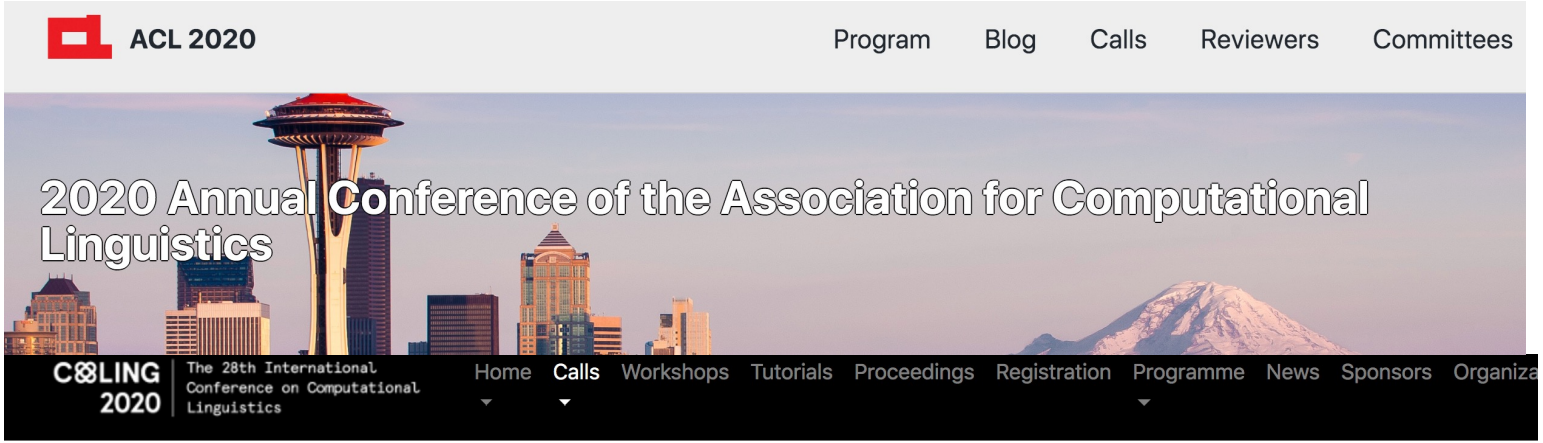
Applications

- Astronomy & Astrophysics
- Automation & Process Control
- Climate / Ecological / Environmental Science
- Customer Relationship Management
- Data Science
- Drug Discovery
- Finance
- Genomics & Bioinformatics
- Healthcare Management
- High Energy Physics
- Intelligence Analysis
- Internet of Things
- Intrusion & Fraud detection
- Logistics Management
- Recommendation
- Risk Management
- Social Network Analysis
- Supply Chain Management
- Other Emerging Applications

Human Factors and Social Issues

- Ethics of Data Mining
- Intellectual Ownership
- Interestingness & Relevance
- Privacy and Fairness Models
- Privacy Preserving Data Mining
- Risk Analysis and Risk Management
- Transparency and Algorithmic Bias
- User Interfaces and Visual Analytics
- Other Human and Social Issues

Conferences and organizations for NLP



COLING'2020 FIRST CALL FOR PAPERS (MAIN CONFERENCE)

The 28 th International Conference on Computational Linguistics (COLING) will be held in Barcelona, Spain, from 13 to 18 September 2020. We invite the submission of papers on original and unpublished research on all aspects of computational linguistics.

Conferences and organizations for NLP

- Applications (including BioNLP and eHealth, NLP for legal purposes, NLP for Social Media and Journalism, among others)
- Dialogue and Interactive Systems
- Discourse and Pragmatics
- Document classification, topic modelling, information retrieval and crosslingual retrieval
- Language Generation
- Language Modelling
- Information Extraction and Text Mining, and Knowledge Graph
- Language Resources and Evaluation
- CL/NLP and Linguistic Theories, Cognitive Modeling and Psycholinguistics
- Machine Learning for CL/NLP
- Multilingual processing, Machine Translation and translation aids
- Morphology and Word Segmentation
- Semantics (of words, sentences, ontologies and lexical semantics)
- Sentiment Analysis, Opinion and Argument Mining
- Speech recognition, text-to-speech and spoken language understanding
- Summarization and Simplification
- Tagging, Chunking, Syntax and Parsing
- Textual Inference and Question Answering
- Vision, Robotics, Multimodal and Grounded Language Acquisition

What did you learn today?

Questions?