Semantic Search & Proactive Discovery
NFAIS 2015, 23 Feb 2015

Alex D Wade
Director, Scholarly Communication
Microsoft Research

@alexwade
Making Cortana the Researcher’s Dream Assistant

Cortana, the personal assistant for Windows Phone 8.1, certainly has received plenty of attention since being announced in April. Her ability to make your day-to-day life easier by managing your information and keeping you up to date has made her one of the most discussed technological developments of the year.

Now, Cortana is about to get smarter—way smarter.

This fall, she will be able to point the way to a wealth of information from the academic community. Cortana is powered by Bing, and soon, that search engine will have academic data tightly integrated and prominently featured on its search pages.

The new functionality, announced July 14 curing the 15th annual Microsoft Research Faculty Summit, represents the latest step for Microsoft Academic Search, which has been a test bed for research ideas for fields such as data mining, named-entity disambiguation, and visualization. This research project has served as a means to explore the challenges in searching for scholarly works, such as achieving author disambiguation, determining publication influence, and constructing graphs of related authors.
Digital Libraries and Autonomous Citation Indexing

Steve Lawrence, C. Lee Giles, Kurt D. Bollacker

The World Wide Web is revolutionizing the way that researchers access scientific information. Articles are increasingly being made available on the homepages of authors or institutions, at journal Web sites, or in online archives. However, scientific information on the Web is largely disorganized. This article introduces the creation of digital libraries incorporating Autonomous Citation Indexing (ACI). ACI autonomously creates citation indices similar to the Science Citation Index. An ACI...

DOI: 10.1109/2.789447

Citation Context (225)

...Many well-known citationindexing systems, such as CiteSeer [1], ISI Web of Knowledge [2] and Google Scholar [3], could have implemented complex reference parsing algorithms, though detailed reports about their algorithms and performance have not been found in the literature...

Xiaoli Zhang, et al. A structural SVN approach for reference parsing

...Lawrence et al., 2002...

Aleksandar Kovačević, et al. Automatic extraction of metadata from scientific publications for CRIS...

...1999) developed a search engine for computer science research papers...
http://academic.research.microsoft.com
Old-MAS: a self-contained Academic graph

Papers:

Derived entities:
- Authors
- Journals
- Conferences
- Organizations
- Field of Study

Enrichment from web:
- Author homepages
- Author Images
- Geolocations
- Crowdsourcing

Academic Graph

Web crawls
Publisher feeds
Open repositories
(Web-scale) Search Engines’ Role in Content Discovery

- ITHAKA longitudinal study shows growing importance of general-purpose search engines among faculty

- Queensland University of Technology research (below) into grad student behavior

- Industry analysis shows 4 billion visits to scholarly publishing sites annually, ~50% from non-affiliated users (search engines as #1 referrer)

- BOTTOM LINE: Billions of users, inside and outside academia, use general-purpose search engines as first path of discovery


http://dx.doi.org/10.1016/j.acalib.2011.04.003
What we hear from students:

"I will put an outline on a Word document, then fill in that outline with information. If there's not enough information that I can find then I'll rethink my thesis"

"I was getting what I searched for but not what I needed"

“When I am researching something that I am already knowledgeable about, I use the search bar, because I would know, when I see it, what I'm looking for. But I don't know about politics, for example, so I feel like I need to see more contextual information in order to understand what I am looking for and how I can better find it."
Web search issues...

- **Imprecise Results**: 1 in 4 delivers successful results.
- **Refinements**: 42% of sessions require refinements.
- **Lengthy Tasks**: 44% of sessions lasting day or more.
Move from Query to Session

• Query is a single action

• Session is a collection of related actions over time
Move from Search Box to Conversations
Bing Knowledge & Cortana
Understanding User Intent

What is the intent of these queries?

- BA48
- Lee Giles
- Weather Berlin
- How many ounces in a litre?
- How tall is the tallest building in the world
- Tallest buildings in the world
- How tall is Barack Obama
- What movies has Ian McKellen been in?
- Italian restaurants
- Books by Cory Doctorow
- Cory Doctorow’s latest book
From Keyword Search ➔ Semantic Search

• Each page on the web is a “bag of words”
• Relationships are HREF links between pages
  • E.g. PageA LinksTo PageB
• Tf-idf ➔ PageRank

• Knowledge of the User
  • Location, Calendar
  • Interests, Preferences, Query History

• Knowledge of the World
  • People, Places, Things, Events
  • Relationships

• A User Experience that is Personal
  • Anticipatory
  • Conversational
  • Proactive
From a graph of *pages* to a graph of *things*
Multi-disciplinary Mash-up

- Distributed Indices
  - Inverted + forward index
- Query federation
- Posting lists + merge sort
- Scalable system best practices

IR / Web Search

- Statistical language model
- Semantic synchronous decoding
- NL generation

Database

- Relational DB knowledge representation
- Row and column indexing

Spoken Language Technology
Toyota

Toyota Motor Corporation is a Japanese automotive manufacturer headquartered in Toyota, Aichi, Japan. In 2013 the multinational corporation consisted of 335,498 employees worldwide and, as of January 2014, is the fourteenth-largest company worldwide.

Founded: Aug 28, 1937
CEO: Akio Toyoda
Founder: Kiichiro Toyoda
Headquarters: Toyota
Place founded: Japan

Stock Price
TM (NYSE) 114.45 ▲ +1.06 (0.93%)
Oct 24, 2024 4:46 PM EDT 20 minutes delayed

Related people
Kiichiro Toyoda
Akio Toyoda
Eiji Toyoda
Shoichiro Toyoda
Katsunori Watanabe

People also search for
Down and Out in the Magic Kingdom is a 2003 science fiction book, the first novel by Canadian author and digital-rights activist Cory Doctorow. Concurrent with its publication by Tor Books, Doctorow released the entire text of the novel under a Creative Commons noncommercial license on his website, allowing the whole text of the book to be read online.

First published: Feb 01, 2003
Author: Cory Doctorow
Genre: Sci-Fi
Awards: Locus Award for Best First Novel

Ratings:
Barnes and Noble
Amazon

Read book
Read for free on Gutenberg
Buy on Barnes and Noble
Buy on Amazon
Seattle Public Library

Read excerpt on OverDrive

People also search for
See all (15+)
From Questions to Conversations
Current weather in District of Columbia

Bing Weather

28°F
Mostly Cloudy
Hi/Lo: 38°/10°

Data from WDT
what is the tallest building in the world

what is the tallest building in seattle
what is the tallest building in the world
what is the tallest tree in the world
what is the tallest mountain in the world
what is the tallest mountain in the usa
what is the tallest building in the us
what is the tallest mountain in north america
what is the tallest mountain in alaska
The world’s tallest man-made structure is the 829.8 m (2,722 ft) tall Burj Khalifa in Dubai, United Arab Emirates. The building gained the official title of "Tallest Building in the World" at its opening on January 4, 2010.

See more on en.wikipedia.org/wiki/List_of_tallest_buildings_and_structures_in_the_world

List of tallest buildings and structures in the world...en.wikipedia.org/wiki/List_of_tallest_buildings_and_structures_in...

This measurement is the most widely used and is used to define the rankings of the 100 Tallest Buildings in the World. Highest Occupied Floor.

Debate over definitions: Tallest structure: Tallest buildings

List of tallest buildings in the world - Wikipedia, the...en.wikipedia.org/wiki/List_of_tallest_buildings_in_the_world


Ranking criteria and - Tallest skyscrapers - Photo gallery

Related searches
Top 10 Tallest Buildings World
Future Tallest Buildings
Saudi Arabia Tallest Building
Tallest Building Dubai
Tallest Building in America 2014
One World Trade Center Completion
Tallest Building in the United States
World’s Tallest Buildings Comparison

Related images
Barack Obama height

6 feet 1 inch (1.85 meters)

Data from IMDb

How tall is Barack Obama? | Height of the Celebrities...
howtallis.info/barack-obama.html
Barack Obama is the 44th and current President of the United States. Obama previously served as the junior United States Senator from Illinois from January 2005 until...

How tall is Barack Obama - Answers.com
www.answers.com › US Presidents › Barack Obama
President Barack Obama is 6' 1¼" (1.87 meters) tall ... She was assigned the role of advisor to Barack Obama who just came from Harvard. They met at a law ...

How tall are Barack and Michelle Obama? What is the...
askville.amazon.com/tall-barack-michelle-obama-average-height...
... How tall are Barack and Michelle Obama? ... Barack Obama will be the next U ... http://www.associatedcontent.com/article/8952/53/how_tall_is_michelle_obama_will ...

How tall is Barack Obama - Celebrity Heights
celebrity-height.com/how-tall-is-barack-obama
Wondring How Tall Is Barack Obama? We have Barack Obama height, most notable career moments all here ... Barack Obama is 6'1" (185cm) How tall is Barack Obama ...

Related searches for how tall is barack obama
How Tall Is George W Bush | How Tall is Kevin Hart
How Tall Is Michelle Obama | How Old is Michelle Obama

Barack Obama
Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he...
Flight status for Alaska 3
MSN Travel 2 minutes ago
DCA → SEA
On time - Departs in 7 hrs 48 min
6:40 PM 9:44 PM
February 23 (Monday) February 23 (Monday)
Washington Seattle
Terminal B N/A
View Airport Map View Airport Map

Alaska Airlines - Official Site
www.alaskaair.com
Alaska Airlines offers low airfare on all airline tickets, including discount flights on Hawaii vacations and Mexico vacation packages. Book your travel today.
Reservations · Check Flight Status · Check-in Options · Web Check-In

Alaska Airlines (AS) #3 FlightAware
flightaware.com/live/flight/ASA3
Alaska Airlines (AS) #3 Flight Tracker (ASA3) Flight Tracker (en route flights, arrivals, departures, history) with live maps and aircraft photos
Guardians of the Galaxy - Cast

Chris Pratt
Peter Quill

Dave Bautista
Drax

Zoe Saldana
Gamora

Michael Rooker
Yondu Udonta

Djimon Hounsou
Korath

Bradley Cooper
Rocket (voice)

Ophelia Lovibond

Gregg Henry

Lee Pace
Ronan

Guardians of the Galaxy - MSN Movies

Guardians of the Galaxy is a 2014 American superhero film based on the Marvel Comics superhero team of the same name, produced by Marvel Studios and distributed by Walt Disney Studios Motion Pictures. It is the tenth installment in the Marvel Cinematic Universe. The film was directed by James Gunn, who wrote the screenplay...
Chris Pratt

Wikipedia, the free encyclopedia

Oct 04, 2014  Christopher Michael "Chris" Pratt (born June 21, 1979) is an American actor. He is known for television roles, including Bright Abbott in Evenwood, and...

Early life  Career  Personal life  Filmography

News about Chris Pratt

3 Unintentionally Genius Pieces of Business Advice From Chris Pratt

Inc.  8 hours ago
The scruffy everyman-turned-action star's summer success story yielded 3

Chris Pratt

Christopher Michael "Chris" Pratt is an American actor. He is known for
Chris Pratt spouse

**Anna Faris**

(m. 2009)

[Data from Wikipedia](https://en.wikipedia.org/wiki/Anna_Faris)

Anna Faris is an American actress and singer. She is known for her comedic roles as Cindy Campbell in the Scary Movie film series, Kelly in Lost in Translation, Shelley in The House Bunny, Serena in Waiting..., Hannah in My Super Ex...+

**Chris Pratt** - IMDb

www.imdb.com/name/nm0695435/bio

Chris grew up in Lake Stevens, WA. He is married to fellow Washington State native, ... Spouse (1) Anna Faris (m. 09). Chris Pratt Personal Details: Biography; 35 years old - News - Awards - Films

Who is actor Chris Pratt married to? - Ask questions, Find ... askville.amazon.com/actor-Chris-Pratt-married/AnswerViewer.do?...>

... Who is actor Chris Pratt married to? ... Spouse(s) Anna Faris ... "Chris Pratt is married to Anna Faris."

**Chris Pratt** - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Chris_Pratt

Chris Pratt; Pratt at the ... Spouse(s) Anna Faris (m. 09) ... They married in Bali on July 9, 2009. In August 2012 their son Jack was born. Filmography ... Early life - Career - Personal life - Filmography
Anna Faris Place Of Birth - Anna Faris Net Worth
www.getnetworth.com/tag/anna-faris-place-of-birth
Anna Faris Place of Birth? Anna Faris Net Worth is $14 Million. Anna Faris is an actress and singer with a net worth of $14 million. Anna Faris earned her net worth ...

Anna Faris biography, birth date, birth place and pictures
www.browsebiography.com/bio-anna-faris.html
Anna Faris is credited as Actress and comedian, Scary Movie series, My Super Ex-Girlfriend. Anna Kay Faris is an American actress, comedian, and singer. She is ...

Anna Faris - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Anna_Faris
Anna Kay Faris (born November 29, 1976) is an American actress and singer. She is known for her comedic roles as Cindy Campbell in the Scary Movie film series (2000 ... Early life · Career · In the media · Personal life · Filmography · Discography

Anna Faris - Biography and facts - WhoIsLog.info
www.whoislog.info · People · F
Anna Kay Faris (ˈənə fərɪs; born November 29, 1976) is an American actress and singer. She is known for her starring role in the Scary Movie film series ...
New MAS: Academic is a subset of larger graph

Authors
Journals
Conferences
Organizations
Field of Study

Derived entities:

Enrichment from web:

Author homepages
Author Images
Geolocations
Crowdsourcing

Academic Graph
Disambiguation
Andrew Ng doctoral advisor

Michael I. Jordan

Data from Wikipedia

Andrew Ng - Executive Bio, Compensation History, and ...
people.equilar.com/bio/andrew-ng-stanford-university/salary/606460
Andrew Ng is an advisor to Lex Machina on machine learning and artificial intelligence. An Associate Professor at Stanford University, Andrew is a member of both the ...

Andrew Ng | Academic Arts High School | ZoomInfo.com
www.zoominfo.com/p/Andrew-Ng/1542725979
Board Member at Academic Arts High School - West St. Paul, Minnesota
View Andrew Ng’s business profile as Board Member at Academic Arts High School and see work ... Andrew Ng is starting his third year as the mathematics advisor at ...

What is it like to have Andrew Ng as your PhD advisor? - Quora
www.quora.com/... Stanford University
What does Andrew Ng's new job at Baidu mean for his involvement at Stanford as a professor?

Andrew Ng - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Andrew_Ng
Andrew Ng, From Wikipedia, the free encyclopedia. Jump to: navigation, search. ... Doctoral advisor: Michael I. Jordan. Known for: Deep Learning, MOOC: Machine learning research · Online education · Personal life

Michael I. Jordan

Michael Irwin Jordan is an American scientist. Professor at the University of California, Berkeley and leading researcher in machine learning and artificial intelligence. Jordan was born in Ponchatoula, Louisiana, to a working-class fa. +

www.oeoa.borutu...
en.wikipedia.org

Born: Feb 25, 1966 (age 58) · Louisiana, United States
Education: Arizona State University · University of California, San Diego
Academic advisor: David Rumelhart

Timeline
1988: He was a professor at MIT from 1988-1998.
See more →
Journals

**Cell**

Cell is a peer-reviewed scientific journal publishing research papers across a broad range of disciplines within the life sciences. Areas covered include molecular biology, cell biology, systems biology, stem cells, developmental biology, etc.

**Applied Physics Letters**

Applied Physics Letters is a weekly peer-reviewed scientific journal that is published by the American Institute of Physics. Its focus is rapid publication and dissemination of new experimental and theoretical papers regarding applications of physics in all disciplines of science, engineering, and modern technology. Additionally, there is an emphasis on fundamental and new developments which lay the groundwork for fields that are rapidly evolving.

**Machine Learning**

Machine Learning is a peer-reviewed scientific journal, published since 1995. In 2001, forty editors and members of the editorial board of Machine Learning resigned in order to found the Journal of Machine Learning Research, saying that in the...
Conferences

KDD 2015
We invite submission of papers describing innovative research on all aspects of knowledge discovery and data mining, ranging from theoretical foundations to novel models and algorithms for data mining problems in science, business, medicine, and engineering. Visionary papers on new and emerging topics are also welcome, as are appl... →

wikicfp.com
Dates: Aug 10 - 13, 2015
Location: Sydney
Subject: Data mining
Website: KDD 2015
Submissions due: Feb 20, 2015


KDD 2014
KDD-2014 will continue the tradition and its technical program will feature the research track, the applications track, the industry practice track, the KDD Cup competition, workshops, tutorials, oral presentations, keynotes, poster sessions... →

www.kdd.org
Location: New York City
Dates: Aug 24 - 27, 2014
Subject: Data mining
Website: KDD 2014
Abstracts due: Feb 13, 2014
Submissions due: Feb 21, 2014

Data from: kdd.org
Feedback

Related searches
KDD Cars
KDD Cup
KDD 2015
KDD Kuwait
KDD Nuggets
KDD 2014 Accepted Papers
KDD Group
Fields of Study
Relationships

Normal distribution
In probability theory, the normal distribution is a very commonly occurring continuous probability distribution—a function that tells the probability that any real observation will fall between any two real limits or real numbers, as the curve approaches... en.wikipedia.org

Abraham de Moivre was a French mathematician known for de Moivre's formula, one of those that link complex numbers and trigonometry, and for his work on the normal distribution and probability theory.

Related
Carl Friedrich... Francis Galton Abraham de Moivre Ronald Fisher Theodore Wilbur Ann...

People also search for See all (10+)
Poisson distribution Binomial distribution Log-normal distribution Exponential distribution Uniform distribution

Black hole
A black hole is a region of spacetime from which gravity prevents anything, including light, from escaping. The theory of general relativity predicts that a sufficiently compact mass will deform spacetime to form a black hole. Around a blac... en.wikipedia.org

People also search for: Wormhole · Supermassive black hole · Neutron star · Spacetime · Event horizon · Supernova · White hole · Galaxy

Among his significant scientific works have been a collaboration with Roger Penrose on gravitational singularity theorems in the framework of general relativity, and the theoretical prediction that black holes emit radiation, often called Stephen Hawking radiation.

See all (10+)
Stephen Hawking Albert Einstein John Michell Subrahmanyan Chandrasekhar Karl Schwarzs...
Anticipatory: Query Auto-suggest

papers by andrew ng about
papers by andrew ng about machine learning
papers by andrew ng about artificial intelligence
papers by andrew ng about pattern recognition
papers by andrew ng about natural language & speech
Proactive Discovery
Interesting things for you at Microsoft Faculty Summit

Events

Opening keynote
Kodiak, 9:00 AM to 10:15 AM
You know the presenter Harry Shum

Frontiers in Speech and Language
St. Helens, 10:15 AM to 11:45 AM
You know the presenter Dong Yu

Presenters you might know

Harry Shum
Co-authored papers together

Dong Yu
Co-authored papers together

HARRY SHUM

about academic

Harry Shum

Heung-Yeung "Harry" Shum is Executive Vice President, Technology & Research at Microsoft. He is known for his research on computer vision and computer graphics, and for the development of web search engine Bing.

en.wikipedia.org

born

1966 | China

current

Executive Vice President

location

Redmond, Washington

Summary

Affiliation: Microsoft
Publications: 405
Citations: 10836
Fields: Graphics, Computer Vision, Multimedia

How you're related

Co-authored Improved name recognition with user modeling
ZHI DING

about
Zhi Ding

Over 30 years of R&D experience in communications and signal processing. Professor for 23 years. Experienced consultant in engineering and as litigation expert. Started recreational running in 2010. Completed 1 marathon and 3 half marathons.

www.zhidin.com

current
Professor at UC Davis

location
Davis, California

ZHI DING

academic
Summary
Affiliation: University of California Davis
Publications: 320
Citations: 2512
Fields: Natural Language & Speech, Networks & Communications, Electrical & Electronic Engineering

ZHI DING

experience
Professor
UC Davis | 2000 - Present
Assoc. Professor of Electrical and Computer
University of Iowa | 1999 - 2000
Consultant
Nortel | 1995 - 2000
More on LinkedIn

ZHI DING

publications
sorted by relevance
 Blind channel identification based on second order cyclostationary statistics
Citations: 40 | Authors: F. Blasch, Z.-D. Ding
For finite impulse response (FIR) channels with uncorrelated input sequences, L. Tong et al. (1999, 1990) have proposed an identification method under the special rank condition...

A novel ultra-wideband pulse design algorithm
Citations: 18 | Authors: J. Zhu, Y. Wang, Z. Ding, K. Wallace
We present a new algorithm to numerically generate pulses that not only have a short time duration for multiple access, but also meet the power spectral constraint of Federal...

ZHI DING

citations
sorted by relevance
Evaluation of biomimetic scaffold of gelatin-hydroxyapatite crosslink as a novel scaffold for tissue engineering:Biocompatibility evaluation with human PDL fibroblasts, human mesenchymal stromal cells, and primary bone cells
Authors: Sahanu Rangayyan, Nikola Dhaene, Sopan Sawdjan, Shi-Yuan Xiong
Biomimetic gelatin (gelatin-hydroxyapatite (HA)) composites have been prepared for studying hard tissue engineering scaffolds. However, the biocompatibility test of this form of material...

Layer-by-layer hyaluronic acid-chitosan coating promotes new collagen ingrowth into a polyethylene terephthalate (PET) artificial ligament in a rabbit model of collateral ligament (MCL) reconstruction model
Authors: Hsiao-Li Ju, Jiang, Yunsheng Gu, Jialing Xu, Pengjun Zhang, Wei Zhang, Ming Chen
The ideal artificial ligament graft should have favorable biocompatibility to facilitate cell adhesion, proliferation, and collagen regeneration. In this present study, surface modification...

ZHI DING

co-authors
S. J. Ben Yoo
University of California Davis
Publications: 433 | Citations: 2707
Fields: Optics & Optoelectronics, Electrical & Electronic Engineering, Networks & Communications
15 co-authored publications

Harvind Samra
University of California Davis
Publications: 433 | Citations: 2707
Fields: Natural Language & Speech, Networks & Communications, Engineering
15 co-authored publications

Soura Dasgupta
Stay tuned for more...

Thank you!