

A Planetary Defense Gateway for Smart Discovery of relevant Information for Decision Support

Myra Bambacus/GSFC & Chaowei Phil Yang/GMU

Ronald Y. Leung, Brent Barbee, Joseph A. Nuth, and Bernie Seery
NASA Goddard Space Flight Center (GSFC)

Yongyao Jiang, Han Qin, Yun Li, Manzhu Yu, Mengchao Xu
George Mason University (GMU)

David S. P. Dearborn
Lawrence Livermore National Laboratory

Catherine Plesko
Los Alamos National Laboratory

- Background
- Framework architecture
- Current results
- Ongoing research
- Conclusions



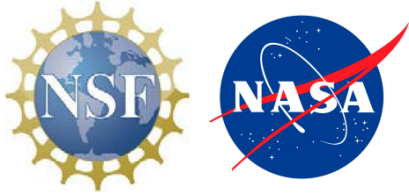
Planetary Defense (PD)



- Near Earth object (NEO) observation
- Design reference asteroids
- Impact modelling
- Decision support
- Mitigation action



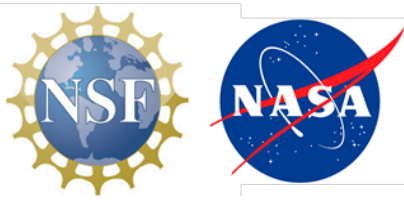
- In this U.S., the NASA Planetary Defense Coordination Office (PDCO) was established in 2016 to study the mitigation of potential Near-Earth Object (NEO) impacts to our home planet.



Motivation for an Information Framework



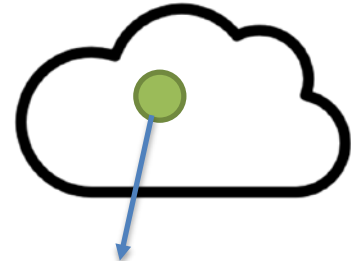
- Information about detecting, characterizing and mitigating NEO threats is dispersed (e.g. publications, briefings.)
- An **overall architecture** to facilitate the collaborations and integrate the different capabilities to achieve the most sensible, executable options for mitigation
- A **cyberinfrastructure** to capture mitigation trades, analyses, model output, risk projections, and mitigation mission design concepts
- Discovery and easy access to knowledge and expert opinion within the project team, as well as factoring in related information from other research and analysis activities



Why Another Resource Discovery Engine?

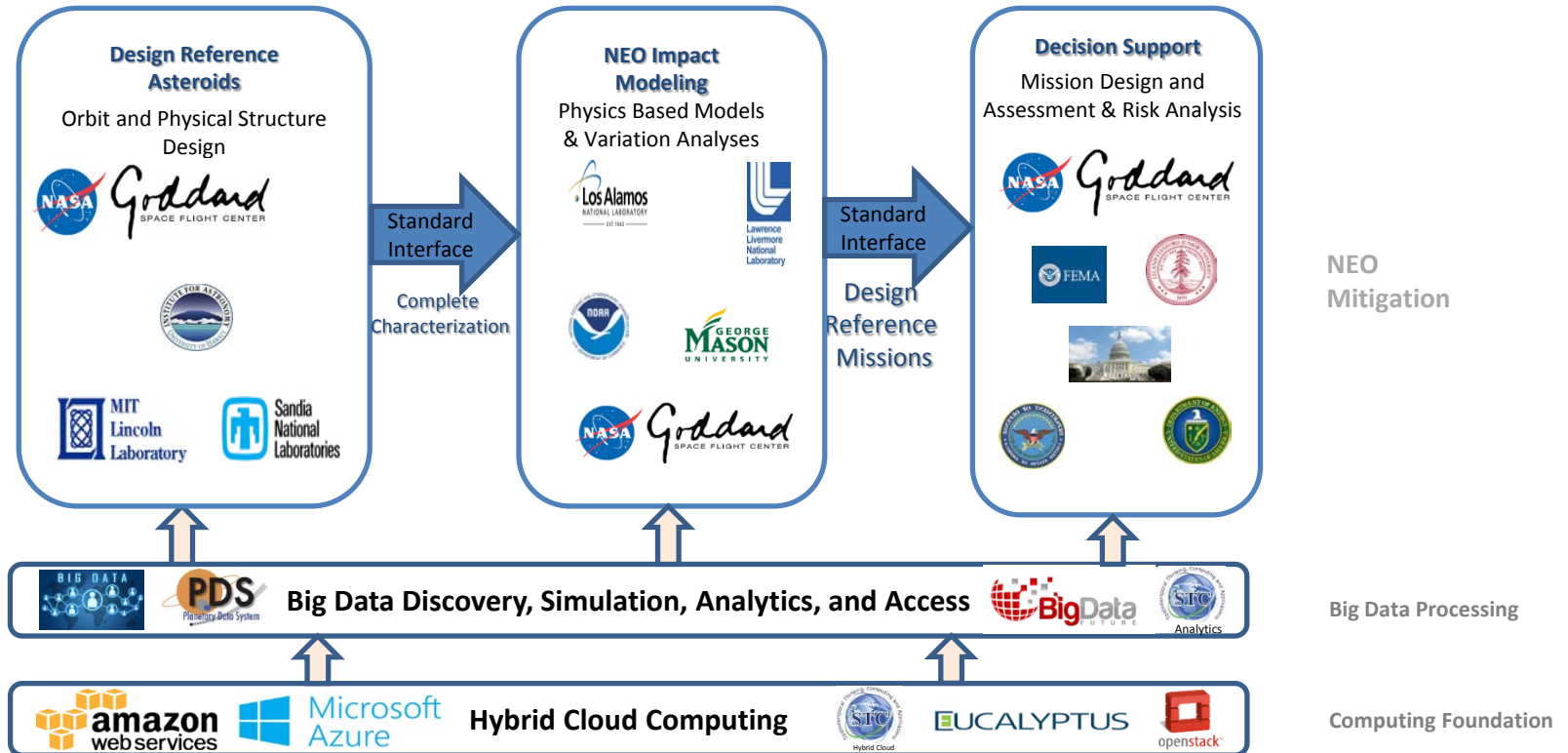


- Domain-specific vs. general-purpose
- Indexed content
 - Google searches from nearly the entire Internet
 - The framework is PD-specific
- Knowledge base
 - Google's *Knowledge Graph* is based on generic sources such as Wikipedia
 - The framework will create a PD ontology aided by domain experts, combined with machine learning and Natural Language Processing (NLP) results
- Decision makers can have easy access to required information and quality knowledge



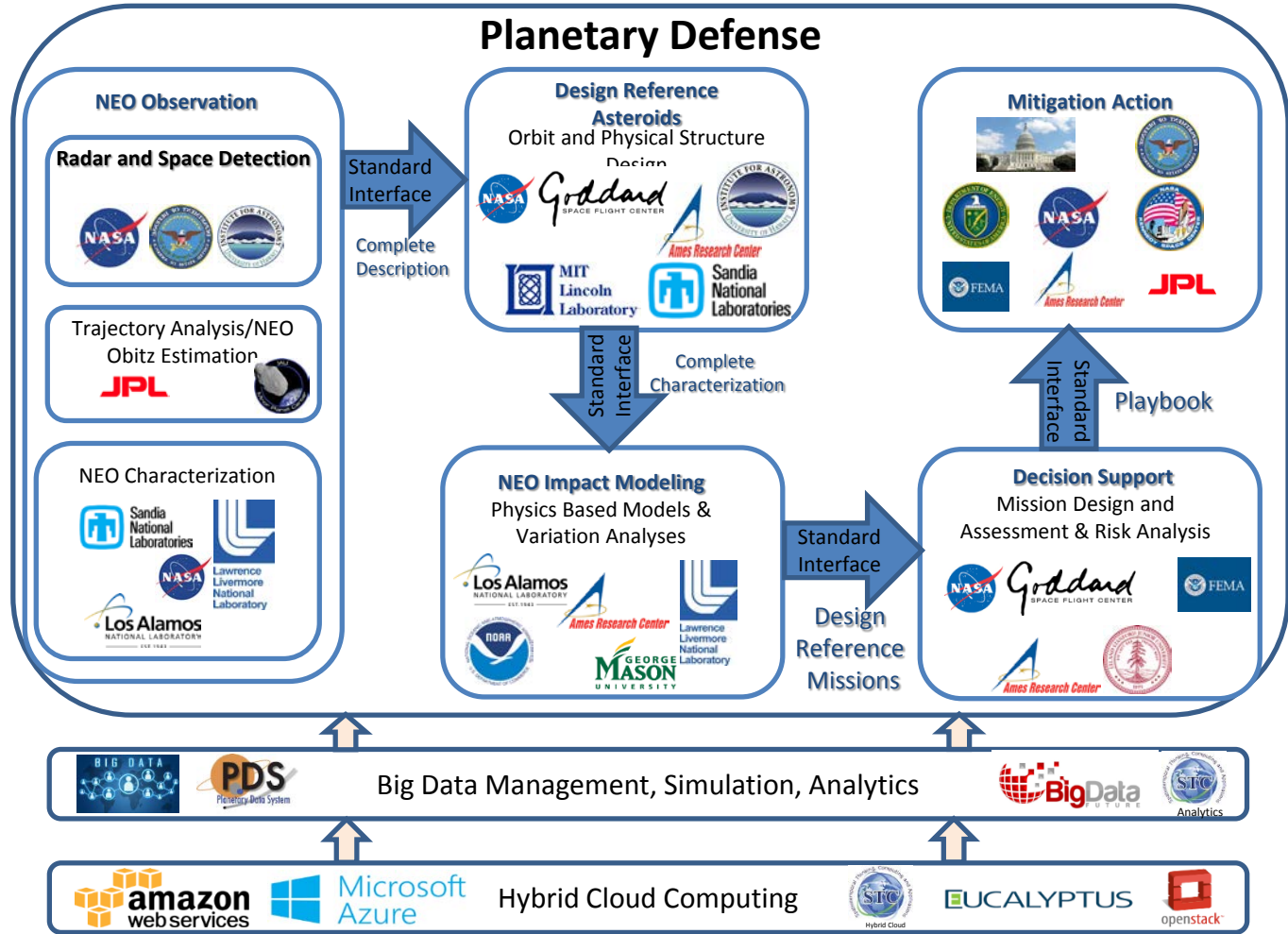
*Planetary Defense
related info*

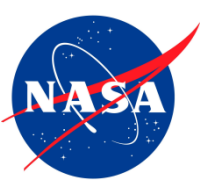
Project Organizational Collaboration



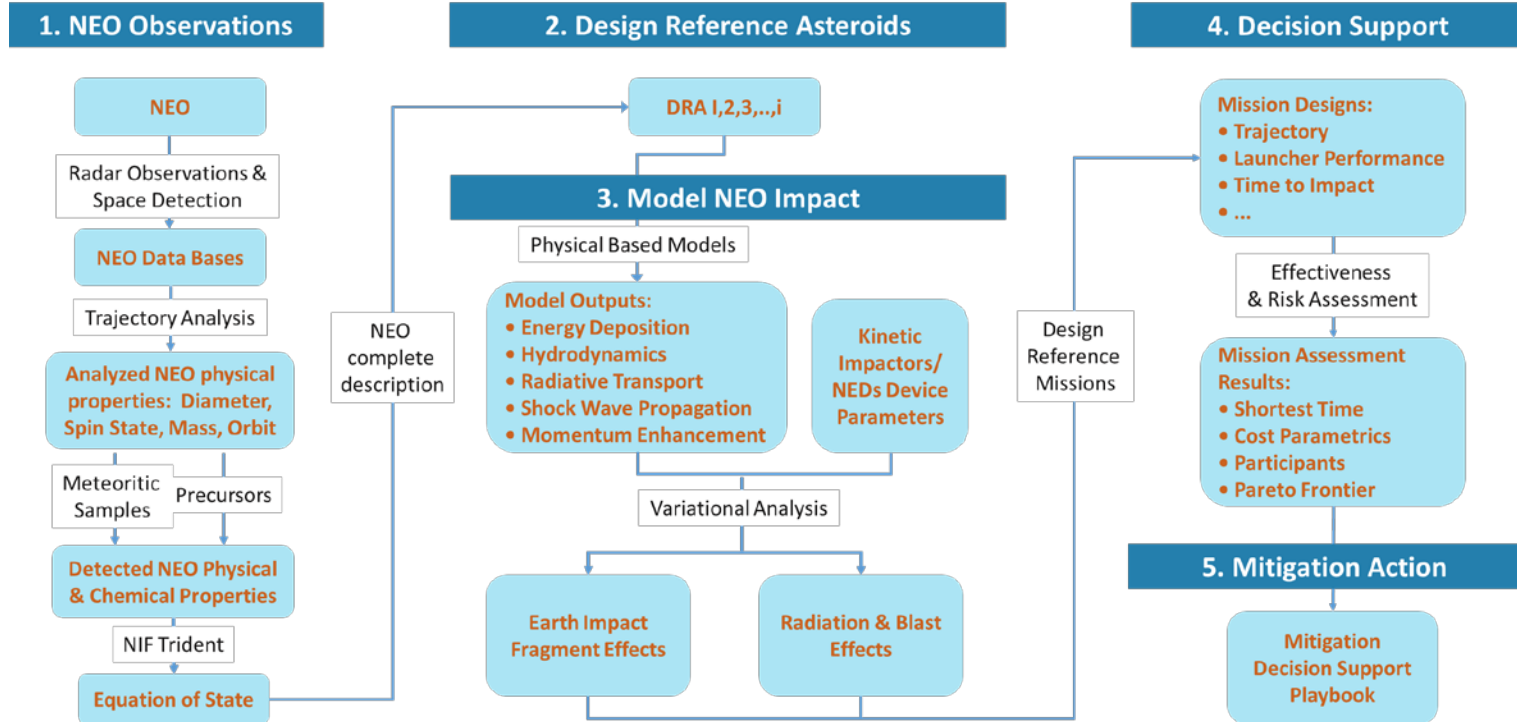
Architectural Framework

List is not exclusive

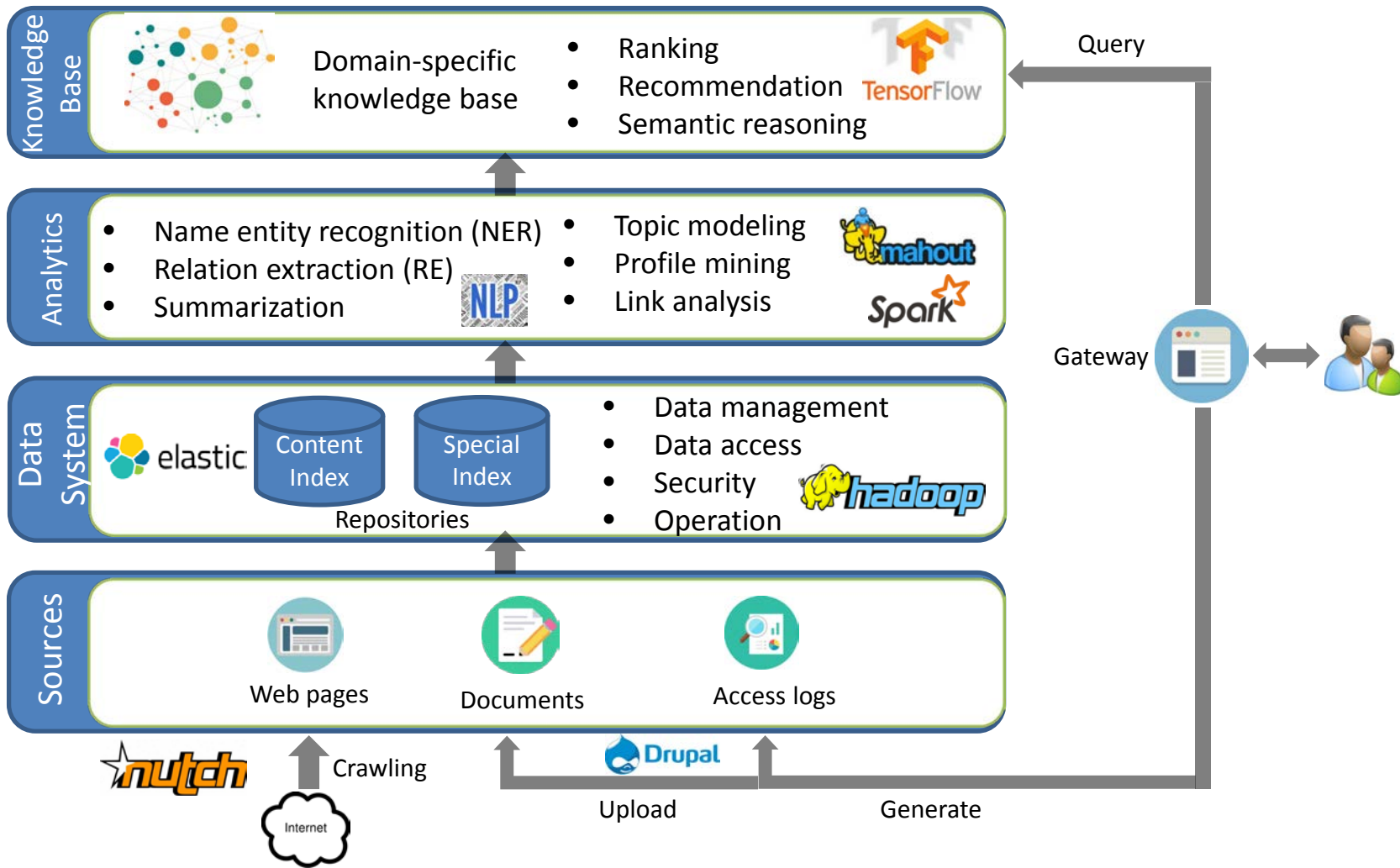


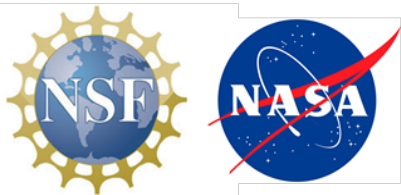


Information Flow



Knowledge Discovery & Usage Framework





Planetary defense (PD) Framework Gateway



- Web Portal: <http://pd.cloud.gmu.edu/>
- User management, document archiving, vocabulary editing
web crawling, search engine

PLANETARY DEFENSE (PD) FRAMEWORK GATEWAY

HOME FILEDEPOT VOCABULARY LIST CRAWLER DB HOW TO USE ACCOUNT INFORMATION

INTRODUCTION

The NASA Planetary Defense Coordination Office (PDCO) was established in 2016 to study the mitigation of potential Near-Earth Object (NEO) impacts to our home planet. NASA Goddard and National Nuclear Security Administration (NNSA) established a collaboration to study the short time response options to potentially hazardous objects (PHOs). The motivation of designing this architectural framework is to develop an integrated architecture for the process of detecting, characterizing and mitigating NEO threats. The project is currently a conglomeration of individual facilities conducting separate research and this framework is meant to help define a collaborative system based on data reporting and sharing across various elements of the collaboration project. With this architecture, the NEO project will increase the efficiency, accuracy, and timeliness of its assessments. The benefits come from—1) maximizing the linkage between different organizations, scientists, and amateurs that are conducting individual NEO research, 2) leveraging the current available computing resources to conduct computation intensive tasks, including data storage, data analysis and modeling processes, 3) developing a sustainable architecture for the NEO defense system— to facilitate faster information sharing, more efficient computing capability, and the creation of a more effective mitigation capability, 4) extending flexibility to future additions or evolutions of assets from the broader PD community. By employing this architecture, the NEO efforts can achieve objectives more efficiently, while further exploring the possible mitigation methods in order to protect our home planet.

The architectural framework is designed to accommodate the process of detecting, characterizing and mitigating Near-Earth Object (NEO) threats. The architecture organizes current data and resources into useful information and correlates that information with the goals of the project. The architectural framework will enable scientists, organizations, and decision makers to locate, identify and resolve definitions, properties, facts, constraints and issues with potentially hazardous asteroids. Our major focus is to design the data and information flow that models the complete process from NEO detection, to the design of mitigation strategies.

The data and information flow depicts what kind of information will be input to and output from each subsystem within the whole process. There are mainly three subsystems: NEO detection, characterization, and mitigation. During the NEO detection process, a diversity of observation systems are utilized to detect the potential NEO. These observations are submitted to databases such as JPL's Minor Planet Center for cataloging. Based on the remote observations, attributes of a NEO are calculated through trajectory analysis, such as diameter, spin state, mass, and orbit. The result of the analysis will trigger NEO characterization process through meteoritic samples or precursor works to obtain the physical and chemical properties. During the NEO characterization process, the calculated NEO properties are fed into physics-based models to calculate the energy deposition, hydrodynamics, and earth impacts, such as earthquakes and tsunamis. The model results are used for mitigation analysis and designing multiple reference missions. Each mission will be assessed for its effectiveness and risks.

The system architecture is designed to describe the supporting infrastructure for the framework. There are four levels of infrastructures that support the mitigation system. First, remote observation systems are needed to detect and track potential NEOs. Second, databases and catalogs are needed for the effective management of NEO records with easy data access and sharing between organizations and agencies. Third, computing infrastructure from Goddard is needed for efficient and accurate calculations for trajectory analysis, effectiveness assessment, risk analysis, and a visualization platform. Finally, HPC clusters at NNSA and NASA are needed to handle the intensive computation during physical based modeling. In addition, the promotion of interoperability, interagency information sharing, and improved computing capabilities can largely facilitate NEO defense decisions. As a strategic information asset, the NEO defense architecture framework will serve as an enabling tool to support the alignment of resources with NEO defense actions, facilitate capability management, and ensure that operational goals and strategies are met.

USER LOGIN

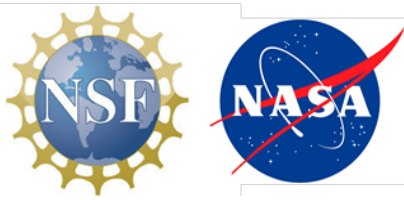
Username *

Password *

[Request new password](#)

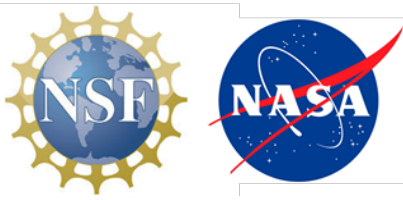
NEO Mitigation

This system is funded by NASA (contract and NSF (WNG14P0001)), and supported by NASA AOT (NNA15AM002). Developed and hosted by NSF Spatiotemporal Innovative Center on the Hybrid Cloud Service.



User management

- User roles: Administer, authenticated user, anonymous user
- Manage access control with permissions and user roles
- Assign permissions and roles to users
- Ban an IP address - The Ban module allows administrators to ban visits to their site from individual or a range of IP addresses.



FileDepot Module: File/document Management



- Create folders or upload new files

FILEDEPOT

New Folder New File more actions... Tags Keyword Search Search

Reports

- Latest Files
- Notifications
- Owned by me
- Downloaded by me
- Unread Files
- Locked by me
- Flagged by me

Recent Folders

- For Brent
- Analysis_Inputs
- Baseline_DART_Spacecraft_M
- Didymos_B_Shape_Model

Top Level Folders

- Case Study 2
- DART
- NASA/NNSA TIM IV
- PD files
- Public

PD files

Filename	Author(s)	Upload date	Action
2014 09 26_.docx	DOEUSER	10/04/16	
Under the auspices of the planned Interagency Agreement between the National Aeronautics and Space Administration and National Nuclear Security Administration, the 2nd Technical Interface Meeting was			
AGU-2015-P_.pptx	cisc	10/04/16	
Enabling Dust Storm Forecasting with Spatiotemporal Optimization and Cloud Computing4. Design Reference Modeling References: Retrieved October 16, 2015, from https://planetary-defense.arc.nasa.gov/wor			
Attach4_.pptx	william.czajkowski	10/04/16	
FY 14 February Monthly Program ReviewDr. Kevin Greenaugh Stockpile Management Overview Nuclear Security Enterprise Overview Office of Research, Development, Test, and Evaluation (NA-11): responsible			
Poster11_.pdf	N/A	10/04/16	
BS_Pos_Asteroid_rev5Short Warning Time Response Options for Hazardous NEOs ???-RLQW?RYHUQPHQW?\$JHQF?7HDP?V?(QG?WR?(QG?0RGHOLQJ?DQG?6LPXODWLRQ?\$FWLYLWLHV Lawrence Livermore National Laboratory Kirs			
For Brent		11/15/16	
Joint NASA-NNSA publications		10/04/16	
NASA publications		10/04/16	



Vocabulary editing module

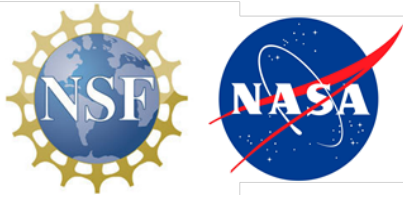


- 130+ concepts
- Create and edit landing page for each concept
- Different user roles have different permissions

VOCABULARY LIST

View Edit

1 <ul style="list-style-type: none">• 101955 Benu (1999 RQ36)	I <ul style="list-style-type: none">• IAU Minor Planet Center• Impact Disaster Planning Advisory Group• Impact experiments• Impactor• Infrasonic Measurement• International Asteroid Warning Network• International Astronomical Union• International Characterization Capability	O <ul style="list-style-type: none">• Object assessment / characterization• Orbit Determination/Estimation• Orbital debris tracking• OSIRIS-REx
3 <ul style="list-style-type: none">• 3-D Visualization	K <ul style="list-style-type: none">• Ka-band system• Kinetic impactor• Knowledge integration• Knowledge Reasoning Models	P <ul style="list-style-type: none">• Perturbation Climatology• Petascale supercomputing• Physics-based modeling• Planetary Defense Information Architecture (PDIA)• Planetary Defense Policy Development• Planetary Defense Strategies• Post Encounter• Post Encounter Termination• Post Launch Check Out• Pre-launch• Precursor mission• Public/citizen engagement
A <ul style="list-style-type: none">• Acid Trauma• Airburst modeling• Approach• Arecibo Observatory• Asteroid class• Atmospheric breakup	L <ul style="list-style-type: none">• Land, Water, and Atmospheric NEO Impact Effects• Launch• Launch segment• Light Curve Database• Luminous efficiency	Q <ul style="list-style-type: none">• Quantitative Risk Metrics
B <ul style="list-style-type: none">• Bistatic Doppler effect• Blast and Thermal Propagation• Bolides	M <ul style="list-style-type: none">• Material Equations of state• Meteorite samples• Meteorite Sampling• Miss• Mitigation• Mitigation Communication• Mitigation Data Assessment / Analysis• Mitigation Data Collection• Mitigation decision / deployment options• Mitigation determination options• Mitigation execution	R <ul style="list-style-type: none">• Radial velocity• Radiative transport• Radio telescopes• Response Time (after warning time)• Return (situational awareness)• Risk Analysis
C <ul style="list-style-type: none">• Casualty Sensitivity• Chelyabinsk• Chemical composition• Chondrite• Command• Communications segment• Contact	N	S <ul style="list-style-type: none">• Safe mode• Secondary Effects• Seismic shaking• Space Segment
D <ul style="list-style-type: none">• Data centralization• Data flow• Data management• Data mining• Data sharing• Decision Support Analysis• Deep space station• Design reference Asteroids		



Web crawling module



- Nutch: Open Source web crawler
- Store them in Elasticsearch (full-text search engine)
- 5 seed URLs
- Similarity between page vs. vocab list
- Baseline

```
controller.addSeed("http://neo.jpl.nasa.gov/");
controller.addSeed("http://global.jaxa.jp/");
controller.addSeed("http://neo.ssa.esa.int/");
controller.addSeed("http://neocam.ipac.caltech.edu/");
controller.addSeed("http://www.minorplanetcenter.net/iau/mpc.html");
```

CRAWLER DB

Enter planet concepts (e.g. Mars) to search

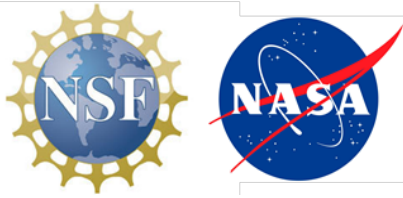
Showing first 100 entries (from 48000 matched entries)

[Near-Earth Object Program](#)
NASA Jet Propulsion Laboratory (JPL)
• View the NASA Portal Small Asteroid Flew Safely Past Earth Today (2016 RB1) September 7, 2016 A small asteroid designated 2016 RB1 safely flew past Earth today at 10:20 a.m. PDT (1:20 p.m. EDT / 17:20 UTC) at a distance of about 25,000 miles (40,000 kilometers, or just less than 1/10th the distance of Earth to the moon). Full Story Small Asteroid Is Earth's Constant Companion (2016 HO3) June 15, 2016 A small asteroid has been discovered in an orbit around the sun that keeps it as a constant... [More]

[News | NASA Office to Coordinate Asteroid Detection, Hazard Mitigation](#)
NASA Jet Propulsion Laboratory (JPL)
NASA Jet Propulsion Laboratory California Institute of Technology Skip Navigation menu and search Jet Propulsion Laboratory California Institute of Technology close menu menu about JPL about JPL executive council history annual reports contact us opportunities public events overview tours lecture series speakers bureau team competitions special events education Intern Learn Teach News Events news latest news press kits fact sheets media information blog missions current past future proposed all ... [More]

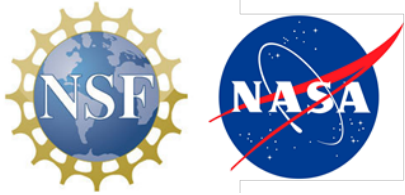
[News | New Details on Ceres Seen in Dawn Images](#)
NASA Jet Propulsion Laboratory (JPL)
NASA Jet Propulsion Laboratory California Institute of Technology Skip Navigation menu and search Jet Propulsion Laboratory California Institute of Technology close menu menu about JPL about JPL executive council history annual reports contact us opportunities public events overview tours lecture series speakers bureau team competitions special events education Intern Learn Teach News Events news latest news press kits fact sheets media information blog missions current past future proposed all ... [More]

[Small Asteroid Is Earth's Constant Companion](#)
NASA Jet Propulsion Laboratory (JPL)
• View the NASA Portal Small Asteroid Is Earth's Constant Companion June 15, 2016 Asteroid 2016 HO3 has an orbit around the sun that keeps it as a constant companion of Earth. Credit NASA/JPL-Caltech A small asteroid has been discovered in an orbit around the sun that keeps it as a constant companion of Earth, and it will remain so for centuries to come. Image of asteroid 2016 HO3 taken on June 10, 2016 by Denise Hung and Dave Tholen of the University of Hawaii using UH's 2.24-meter telescope... [More]



Ongoing research

- Domain specific crawling
- Knowledge extraction from plain text



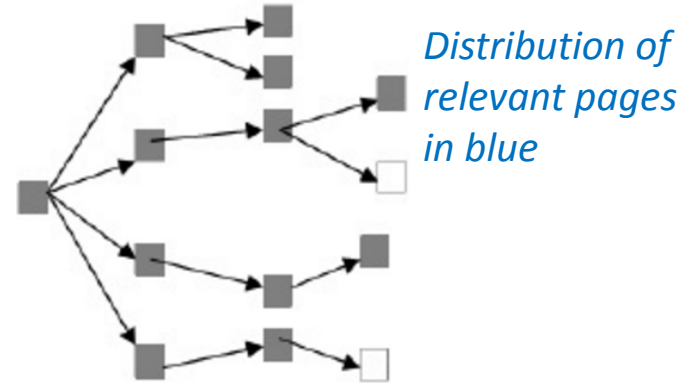
Domain specific crawling

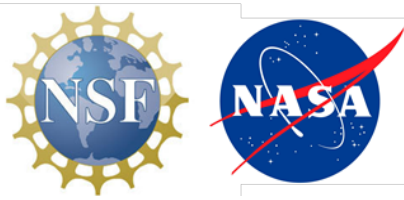


Simplest approach: filter web pages using a **keyword list** (e.g. NEO, asteroid, Bennu, ...) composed by domain experts.

Problems:

- Expensive
- Difficult to exhaust
- Difficult to assign weights to different keywords
- Treat all web pages equally (a page on NASA website and a random one)



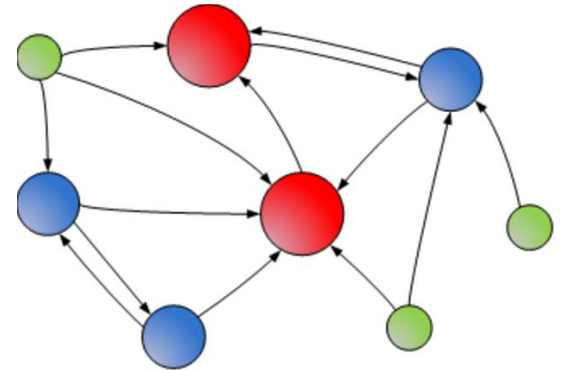


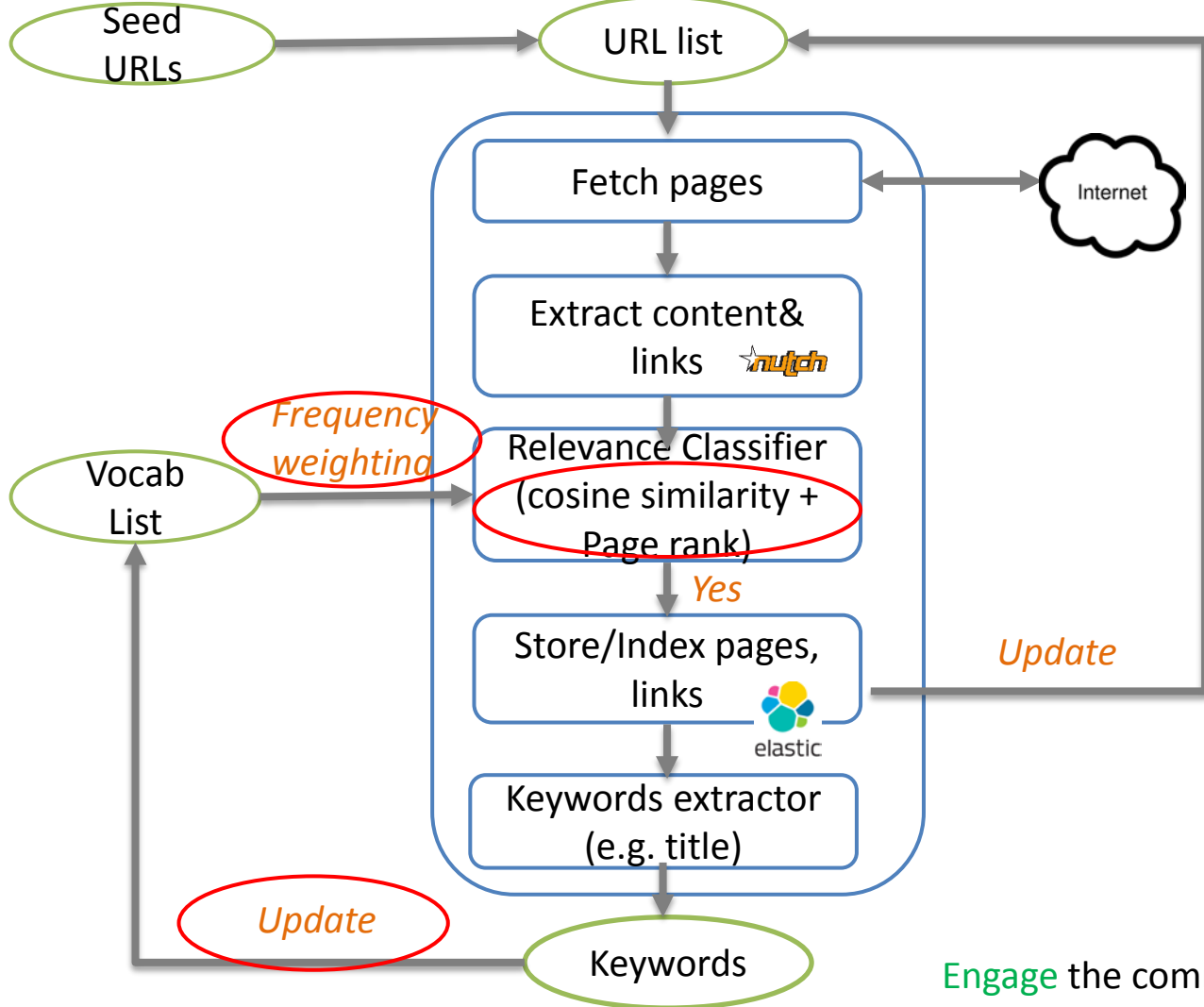
Domain specific crawling



Existing tools in Open Source crawler (e.g. Nutch):

- Link-based
 - Scoring links (OPIC, PageRank scoring)
 - Breadth first or Depth first crawl
- Content-based
 - URL, mimetype filter
 - **Cosine Similarity scoring filter (what we are using)**
 - Naive Bayes parse filter

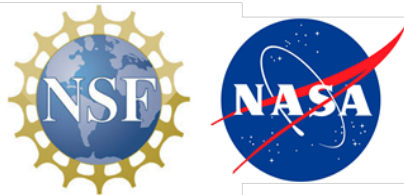




Proposed method

- **Combine** content and link-based scoring to boost the authoritative and relevant web pages
- **Dynamically** update/grow the vocab using info (e.g. title) from the web pages
- **Weight** keywords based on **frequency clustering** (i.e. more frequently seen terms have more weights)

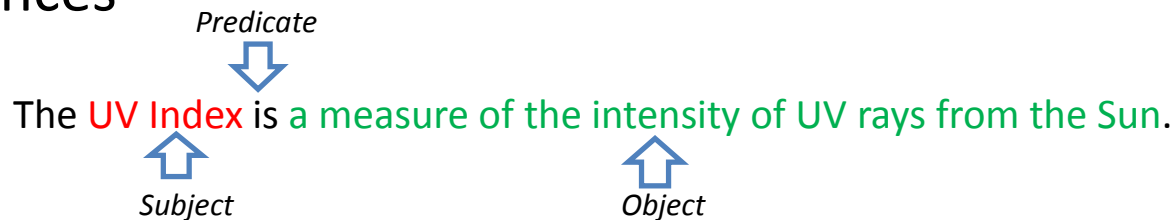
Engage the community to help with the evaluation



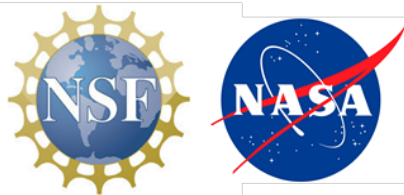
Knowledge extraction from plain text



- Goal: Extract structured information from unstructured web pages and user uploaded documents
- **Relation extraction** in NLP: finding semantic triples (*SPO*) from sentences



- Pattern-based, supervised, semi-supervised, and open information extraction

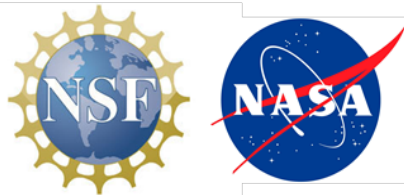


Relation extraction



Hand-written patterns

- “Y such as X”
- “such Y as X”
- “X or other Y”
- “Y including X”
- + Tend to be high-precision
- + Tailored to specific domains
- - Human patterns are often low-recall
- - Hard to be exhaustive



Open Information Extraction



- Recently published by Univ. of Washington
- Extract relations from the sentences with no training data, no list of relations (unsupervised)
- Self-learning process, syntactic and lexical/semantic patterns

The U.S. president Barack Obama gave his speech on Tuesday to thousands of people.

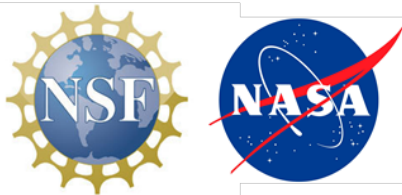


(Barack Obama, is the president of, the U.S.)

(Barack Obama, gave, his speech)

(Barack Obama, gave his speech, on Tuesday)

(Barack Obama, gave his speech, to thousands of people)

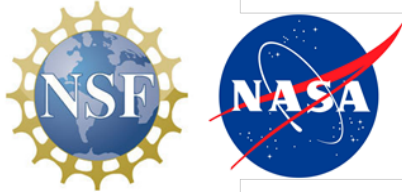


Open Information Extraction



▶ the GRSST	is	a truly international project with over \$18 Million US Today
▶ Jason-1	has	a repeat period of approximately 10 days with 254 passes per cycle
▶ Jason-3	is	capable of measuring significant wave height, sigma naught (σ_0), dry and wet tropos
▶ The Aquarius instrument	has	3 radiometer beams in push-broom alignment with footprint resolutions of 76 km
▶ Jason-3	has	a repeat period of approximately 10 days with 254 passes per cycle
▶ Jason-1	is	capable of measuring significant wave height, σ_0 , dry and wet troposphere and ionos
▶ Level-2 data	refer	to monthly estimates of spherical harmonic coefficients of the Earth gravity field
▶ no downlink signal	was detected	At the beginning of the next contact at 0249 UTC
▶ sensors	included	a CTD at the near-surface and another at 6 m depthFor SPURS-1

- Some are reasonable, some are noise
- Working on reducing noise/identifying reasonable results



Conclusion and Next Steps



- The proposed architecture framework benefits the PD community by
 - Providing discovery and easy access to the knowledge and expert opinion within the project team
 - Maximizing the linkage between different organizations, scientists, engineers, decision makers, and citizens
- Next steps
 - Develop a knowledge base & search ranking for NEO mitigation resources
 - Investigate a knowledge reasoning model for potential mitigation by assimilating existing scenarios
 - Build a 4D visualization tool based on new datasets and existing tools

References

- Agichtein, E., Brill, E. and Dumais, S., 2006. Improving web search ranking by incorporating user behavior information. ed. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 19-26.
- AlJadda, K., et al., 2014. Crowdsourced query augmentation through semantic discovery of domain-specific jargon. ed. *Big Data (Big Data), 2014 IEEE International Conference on*, 808-815.
- Auer, S., et al., 2007. Dbpedia: A nucleus for a web of open data. *The semantic web*. Springer, 722-735.
- Bach, N. and Badaskar, S. 2007. A survey on relation extraction. *Language Technologies Institute, Carnegie Mellon University*.
- Banko, M., et al., 2007. Open Information Extraction from the Web. ed. *IJCAI*, 2670-2676.
- Bargellini, P., et al., 2013. Big Data from Space: Event Report. European Space Agency Publication.
- Battle, R. and Kolas, D. 2012. Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*, 3(4), 355-370.
- Cook, S., et al. 2011. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS one*, 6(8), e23610.
- Egenhofer, M. J., 2002. Toward the semantic geospatial web. ed. *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*, 1-4.
- Graybeal, J., 2015. *Community Ontology Repository Prototype: Development* [online]. ESIP. Available from: http://testbed.esipfed.org/cor_prototype [Accessed 12/5 2016].
- Jiang, Y., et al. 2016a. A Comprehensive Approach to Discovering and Determining the Semantic Relationship among Geospatial Vocabularies - An Example with Oceanographic Data Discovery. *International Journal of Geographical Information Science*.
- Jiang, Y., et al. 2016b. Reconstructing Sessions from Data Discovery and Access Logs to Build a Semantic Knowledge Base for Improving Data Discovery. *ISPRS International Journal of Geo-Information*, 5(5), 54.
- Joachims, T., 2002. Optimizing search engines using clickthrough data. ed. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 133-142.
- Krisnadhi, A., et al., 2015. The GeoLink modular oceanography ontology. ed. *International Semantic Web Conference*, 301-309.
- Lee, J.-G. and Kang, M. 2015. Geospatial big data: challenges and opportunities. *Big Data Research*, 2(2), 74-81.
- Li, W., Goodchild, M. F. and Raskin, R. 2014. Towards geospatial semantic search: exploiting latent semantic relations in geospatial data. *International Journal of Digital Earth*, 7(1), 17-37.

- Manning, C. D., et al., 2014. The Stanford CoreNLP Natural Language Processing Toolkit. ed. *ACL (System Demonstrations)*, 55-60.
- Miller, G. A. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Nativi, S., et al. 2015. Big data challenges in building the global earth observation system of systems. *Environmental Modelling & Software*, 68, 1-26.
- Noy, N. F. and Musen, M. A., 2000. Algorithm and tool for automated ontology merging and alignment. ed. *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00)*. Available as SMI technical report SMI-2000-0831.
- Pouchard, L., 2013. *ESIP Semantic Portal* [online]. ESIP. Available from: <http://testbed.esipfed.org/node/1243> [Accessed 12/5 2016].
- Raskin, R. and Pan, M., 2003. Semantic web for earth and environmental terminology (sweet). ed. *Proc. of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*.
- Singhal, A. 2012. Introducing the knowledge graph: things, not strings. *Official google blog*.
- Soderland, S., et al. 2010. Adapting open information extraction to domain-specific relations. *AI magazine*, 31(3), 93-102.
- Srivastava, J., et al. 2000. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1(2), 12-23.
- Sun, A., Grishman, R. and Sekine, S., 2011. Semi-supervised relation extraction with large-scale word clustering. ed. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 521-529.
- Wu, L. and Brynjolfsson, E. 2013. The future of prediction: How Google searches foreshadow housing prices and sales. Available at SSRN 2022293.
- Yang C., H. Q., Li Z., Liu K., Hu F., 2016 2016a. Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*.
- Yang C., Y. M., Hu F., Jiang Y., Li Y. 2016b. Utilizing Cloud Computing to Address Big Geospatial Data Challenges. *Computers, Environment, and Urban Systems*.
- Zelenko, D., Aone, C. and Richardella, A. 2003. Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb), 1083-1106.
- Zhou, Y., et al., 2008. Large-scale parallel collaborative filtering for the netflix prize. *Algorithmic Aspects in Information and Management*. Springer, 337-348.