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
Project Organizational Collaboration

Design Reference

Asteroids

Orbit and Physical Structure Design

NEO Impact

Modeling

Physics Based Models & Variation Analyses

Decision Support

Mission Design and Assessment & Risk Analysis

Standard Interface

Complete Characterization

Standard Interface

Design Reference Missions

Big Data Discovery, Simulation, Analytics, and Access

Hybrid Cloud Computing

Big Data Processing

Computing Foundation

NEO Mitigation
Planetary Defense

Architectural Framework

List is not exclusive

NEO Observation
  Radar and Space Detection
    NASA
    RAND Corporation
  Trajectory Analysis/NEO Obitz Estimation
    JPL

NEO Characterization
  Sandia National Laboratories
  Los Alamos National Laboratory
  Lawrence Livermore National Laboratory

Design Reference
  Asteroids
    Orbit and Physical Structure
    Design
    MIT Lincoln Laboratory
    Sandia National Laboratories
  NEO Impact Modeling
    Physics Based Models & Variation Analyses
  NEO Observation
    Standard Interface
    Complete Description
    Characterization

Mitigation Action
  Decision Support
    Mission Design and Assessment & Risk Analysis
    Design Reference Missions
    Standard Interface
    Playbook

NEO Impact Modeling
  Physics Based Models & Variation Analyses

Decision Support
  Standard Interface
  Mitigation Action

Big Data Management, Simulation, Analytics

Hybrid Cloud Computing

Amazon Web Services
Microsoft Azure
Eucalyptus
OpenStack
Information Flow

1. NEO Observations
   - NEO
     - Radar Observations & Space Detection
   - NEO Data Bases
     - Trajectory Analysis
   - Analyzed NEO physical properties: Diameter, Spin State, Mass, Orbit
   - Meteoritic Samples
   - Precursors
   - Detected NEO Physical & Chemical Properties
   - NIF Trident
   - Equation of State

2. Design Reference Asteroids
   - DRA 1, 2, 3, ..., j
   - Physical Based Models
     - Model Outputs:
       - Energy Deposition
       - Hydrodynamics
       - Radiative Transport
       - Shock Wave Propagation
       - Momentum Enhancement
     - Kinetic Impactors/NEDs Device Parameters
     - Variational Analysis
     - Earth Impact Fragment Effects
     - Radiation & Blast Effects

3. Model NEO Impact
   - NEO complete description

4. Decision Support
   - Mission Designs:
     - Trajectory
     - Launcher Performance
     - Time to Impact
     - ...
   - Effectiveness & Risk Assessment
     - Mission Assessment Results:
       - Shortest Time
       - Cost Parametrics
       - Participants
       - Pareto Frontier

5. Mitigation Action
   - Mitigation Decision Support Playbook
Knowledge Discovery & Usage Framework

Knowledge Base
- Domain-specific knowledge base
- TensorFlow

Analytics
- Name entity recognition (NER)
- Relation extraction (RE)
- Summarization
- NLP

Data System
- Content Index
- Special Index
- Repositories
- Hadoop

Sources
- Web pages
- Documents
- Access logs
- Nutch

Repositories
- Data management
- Data access
- Security
- Operation

Query

Gateway

Upload
Generate

Crawling

Internet
Planetary defense (PD) Framework Gateway

- User management, document archiving, vocabulary editing, web crawling, search engine
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
- More actions:
  - Set permissions of specific folder for different user roles
  - Text mining will be performed to find the relations between docs & keywords
Vocabulary editing module

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

Image source: https://en.wikipedia.org/wiki/PageRank
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

  - **Predicate**: The UV Index is a measure of the intensity of UV rays from the Sun.

  - **Subject**:  
  - **Object**:  

- **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 GHRSSST is a truly international project with over $18$ Million USD.
- 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 ($\text{sigma0}$), dry and wet troposphere.
- 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, sigma0, dry and wet troposphere and ionosphere.
- 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 depth for 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


Wu, L. and Brynjolfsson, E. 2013. The future of prediction: How Google searches foreshadow housing prices and sales. *Available at SSRN 2022293*.


