7) MARJORIE HLAVA

Using Linked Data and taxonomies to create a quick-start "smart" thesaurus

1. About the Case Organization

The two current applications of this approach are a large scientific publisher and a large association with a robust publishing platform.

In the first instance, the publisher has a growing, broadly based collection of over 100,000 articles from over 250,000 authors. The second is an association with a defined topical area; its 220,000 articles are from 900,000 authors and 29,000 institutions worldwide.

2. About the Challenge

Thesaurus applications have been in use for many years. With the increasingly complex and interconnected world available through digitization and the internet, the expected connectivity of information is growing at a logarithmic rate—everyone finds themselves buried in huge amounts of information. This has led to an ubiquitous need to track, find, and programmatically tag the information in meaningful ways for enhanced retrieval.

Increasingly, researchers are going directly to Google rather than to publisher websites to fulfill their research needs. This has a direct impact on the already struggling scholarly publishing industry.

Two instances of this challenge have led us to design the Smart Thesaurus. We have already proven the ROI of enriching content with subject metadata; further, we know that a thesaurus—with all of its relationships—provides a sound conceptual platform for information organization and retrieval. The next issue was how to enlarge that thinking to an actual working application for broad conceptual areas. We considered ontologies and other options for searching. We decided to build a proof-of-concept based on enriching and interlinking the data itself—combining Linked Data with a concept thesaurus.

3. What We Did

We started by enriching the data using standard thesaurus (taxonomy term) tagging with our Data Harmony suite of tools (Thesaurus Master, a vocabulary management tool, and M.A.I, our Machine Aided Indexer) to automatically apply the terms to the articles. Next, we used the M.A.I. to tag the data full text inline. That is: wherever a thesaurus concept was mentioned in the full text, triggered by the M.A.I. Concept Extractor Engine, we inserted the taxonomy term as a full XML string, effectively enriching the content directly in the full text of each article.

In parallel, we added a field for Linked Data—to hold URI to a persistent source, such as dbpedia or Wikidata, for each thesaurus term—to the Thesaurus Master application. Therefore, each preferred term can have a link to one or more external resources, from which we can query off definitions, links to other data sources, and other information to create dynamic web interfaces.

We adhered to the ANSI/NISO z39.19 Thesaurus Standard, which is comparable with the ISO 25964-1 Controlled Vocabulary Standard, for all thesaurus/taxonomy term record creation. We used DBpedia as a link to external resources and definitions where applicable.

4. Challenges and Lessons Learned

One problem we encountered is that most Linked Data sources are not granular enough to supply a link for every term in a highly specialized scientific thesaurus; the more specific the term, the less likely that a relevant Linked Data source is available.

The optimal solution to this problem is to create the missing Linked Data pages (on dbpedia, Wikidata, or whatever Linked Data source is in use) to enrich the source data in the publisher's specialized area; further, backlinks to the publisher's topic pages (or other relevant areas of their websites) should be added as external resources. In this way, the organization is contributing to and enriching the growing network of linked data sources available on the web as well as promoting themselves as thought leaders in the industry.

5. Impact and Benefits

The results are dramatic increase in uptake of the articles and interlinking of the core content as well as building the community for each topical area.

The benefits are clear.

- Staff time is saved by better information retrieval, freeing them for other activities
- Better search results for the end user (in this case, researchers)

- Portals with embedded Linked Data can stream dynamically generated content from external data sources (other websites, social media, news, images) alongside the publishers' own content, establishing these portals as "one-stop shops" for researchers—this helps to make the sites "sticky" by keeping users from leaving to search other sites
- Topic pages on publishers' websites enhanced with definitions queried from Linked Data sources without staff curation of content

Without the core thesaurus and the automatic linking this set of tasks would have taken several man years of labor to produce the same result. This solution, with the ever increasing burden of so much content, is scalable.

6. Next Steps

After combining content stored in a highly structured XML markup language (such as JATS), a well-formed topical thesaurus, and Linked Data, the next logical step is to implement a robust RDF triples database underneath the content and other data structures. This will help to make Smart Content—websites that are not only enhanced for human users, but completely machine-readable and able to draw inferences between concepts, objects, and entities elsewhere on the web.

8) NOR AZLINAYATI ABDUL MANAF

Padipedia: building a knowledgebase on paddy production using semantic web technology for MARDI (Malaysian Ministry of Agriculture)

1. About the Case Organization

The Malaysian Agricultural Research and Development Institute (MARDI) is a leading agricultural research centre in Malaysia. MARDI was established in 1969 with the main objectives of generating and promoting new, appropriate and efficient technologies towards the advancement of the agriculture, food and agro-based industries. It is located in Serdang, Selangor and has 32 branches nationwide. MARDI is mandated to conduct research in the fields of science, technical, economy, and social with regards to production, utilization and processing of all crops (except rubber, oil palm and cocoa), livestock and food.

2. About the Challenge

Main objective of the effort

To preserve knowledge about research and development on paddy and make it accessible for knowledge discovery.

Issue or problem

The ageing workforce represents an issue with knowledge loss as retirement occurs. Knowledge loss will threaten the paddy and rice research process and activities in order to improve/sustain the rice industry in Malaysia. Electronic and written memos, journals and books provide expert knowledge but they do not connect the knowledge of the whole value chain of paddy research. Thus the paddy ontology is built to be the repository for storing and connecting the knowledge.

Prior to PADIPEDIA implementation, various divisions that are distributed nationwide were "disconnected" from involvement in the paddy research and development activities, making it more difficult to have an integrated knowledge repository.

3. What We Did

The PADIPEDIA application is a web-based portal with searching capabilities, basic analysis and reporting function using Semantic Technology. The development of paddy ontology is an attempt to utilize semantic web technology for organizing knowledge. The paddy ontology will be the repository for storing and connecting the knowledge of the whole value chain of paddy production which includes breeding, agronomy, production system, pest and disease management, post-harvest and product development which is in line with the research scope of MARDI. Padipedia culminated from the need to preserve paddy knowledge and make it accessible for knowledge discovery. Users can obtain paddy-related information not only from within MARDI but from external sites as well.

Innovations about this effort are:

- Ability to perform Semantic Search based on concepts rather than keywords.
- Ability to perform question answering using natural language.
- Provides user access to published knowledge base in the form of open linked data which goes beyond PADIPEDIA knowledge base by linking users to other publicly available knowledge repositories in the world within the Open Linked Data Cloud.

IKO Conference 2015 - Singapore

Using Linked Data and Taxonomies to Create a *Smart* Thesaurus

Marjorie M K Hlava,
President,
Access Innovations, Inc.
Albuquerque, NM USA

Access Innovations Inc.® maker of Data Harmony taxonomy software

Data Harmony from Access Innovations, Inc.

Points to Linked Data

- Point to relevant resources via URLs
- Leverage the thesaurus for rich ontology
- Link to other data repositories
  - Databases
  - People nets
  - DBpedia
  - External websites
  - Twitter feeds
  - Google News alerts
  - Other resource files

Wikip

Create account Log in

WIKIPEDIA
The Free Encyclopedia

Article Talk Read Edit View history Search

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file

Optics

From Wikipedia, the free encyclopedia

This article is about the branch of physics. For the book by Sir Isaac Newton, see Opticks. For the musical artist, see Optical (artist). For other uses, see Optic (disambiguation).

Optics is the branch of physics which involves the behaviour and properties of light, including its interactions with matter and the construction of instruments that use or detect it.[1] Optics usually describes the behaviour of visible, ultraviolet, and infrared light. Because light is an electromagnetic wave, other forms of electromagnetic radiation such as X-rays, microwaves, and radio waves exhibit similar properties.[1]

Most optical phenomena can be accounted for using the classical electromagnetic description of light. Complete electromagnetic descriptions of light are, however, often difficult to apply in practice. Practical optics is usually done using simplified models. The most common of these, geometric optics, treats light as a collection of rays that travel in straight lines and bend when they pass through or reflect from surfaces. Physical optics is a more comprehensive model of light, which includes wave effects such as diffraction and interference that cannot be accounted for in geometric optics. Historically, the ray-based model of light was developed first, followed by the wave model of light. Progress in electromagnetic theory in the 19th century led to the discovery that light waves were in fact electromagnetic radiation.

Some phenomena depend on the fact that light has both wave-like and particle-like properties. Explanation of these effects requires quantum mechanics. When considering light's particle-like properties, the light is modelled as a collection of particles called "photons". Quantum optics deals with the application of quantum mechanics to optical systems.

Optical science is relevant to and studied in many related disciplines including astronomy, various engineering fields, photography, and medicine (particularly ophthalmology and optometry). Practical applications of optics are found in a variety of technologies and everyday objects, including mirrors, lenses, telescopes, microscopes, lasers, and fibre optics.

Optics includes study of dispersion of light.

About: Optics

An Entity of Type : Thing, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

DBpedia

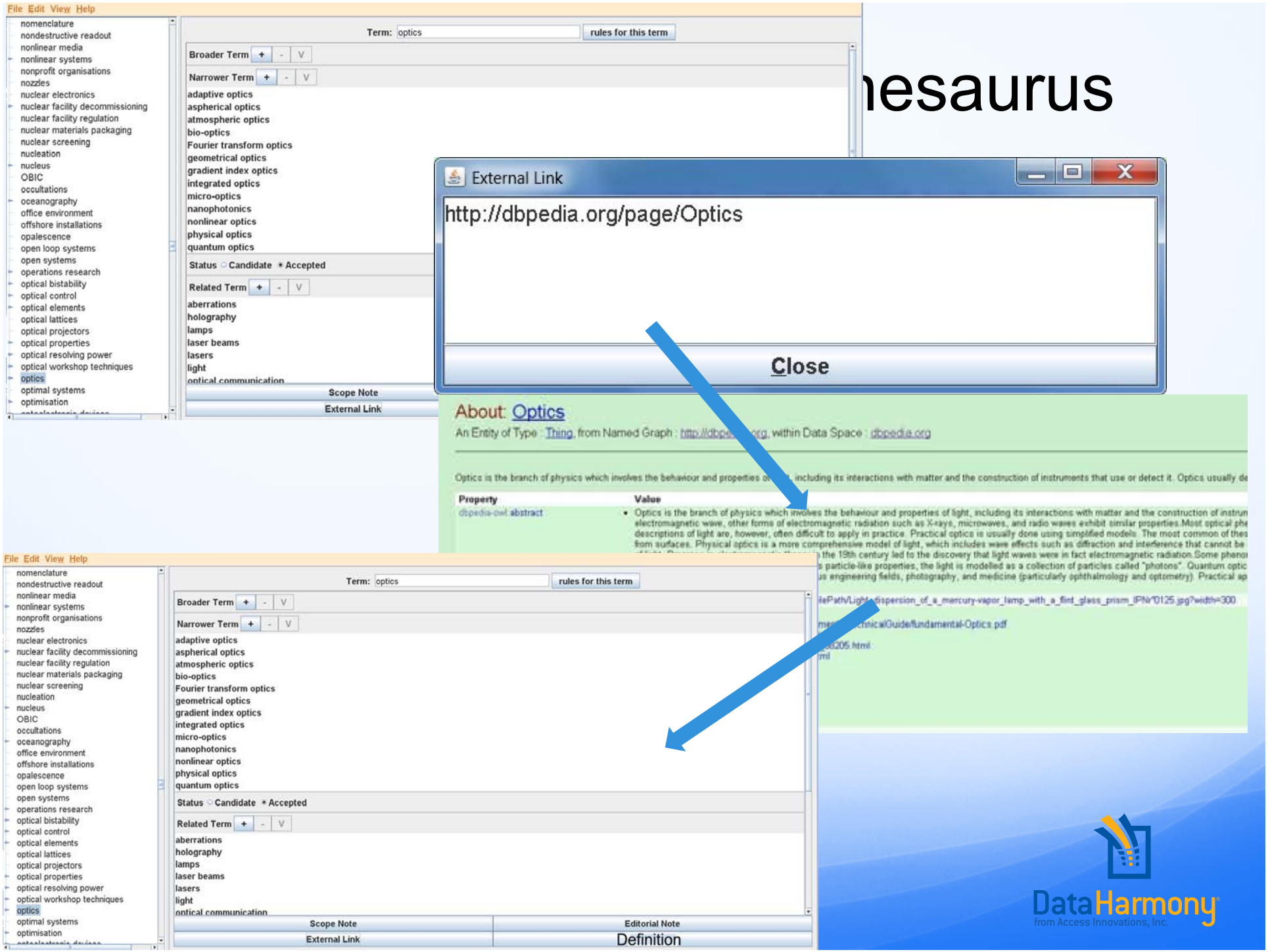
Optics is the branch of physics which involves the behaviour and properties of light, including its interactions with matter and the construction of instruments that use or detect it. Optics usually describes the behaviour of visible, ultraviolet, and infrared light.

Property	Value
dbpedia-owl:abstract	• Optics is the branch of physics which involves the behaviour and properties of light, including its interactions with matter and the construction of instruments that use or detect it. Optics usually describes the behaviour of visible, ultraviolet, and infrared light. Because light is an electromagnetic wave, other forms of electromagnetic radiation such as X-rays, microwaves, and radio waves exhibit similar properties.Most optical phenomena can be accounted for using the classical electromagnetic description of light. Complete electromagnetic descriptions of light are, however, often difficult to apply in practice. Practical optics is usually done using simplified models. The most common of these, geometric optics, treats light as a collection of rays that travel in straight lines and bend when they pass through or reflect from surfaces. Physical optics is a more comprehensive model of light, which includes wave effects such as diffraction and interference that cannot be accounted for in geometric optics. Historically, the ray-based model of light was developed first, followed by the wave model of light. Progress in electromagnetic theory in the 19th century led to the discovery that light waves were in fact electromagnetic radiation.Some phenomena depend on the fact that light has both wave-like and particle-like properties. Explanation of these effects requires quantum mechanics. When considering light's particle-like properties, the light is modelled as a collection of particles called "photons". Quantum optics deals with the application of quantum mechanics to optical systems.Optical science is relevant to and studied in many related disciplines including astronomy, various engineering fields, photography, and medicine (particularly ophthalmology and optometry). Practical applications of optics are found in a variety of technologies and everyday objects, including mirrors, lenses, telescopes, microscopes, lasers, and fibre optics.
dbpedia-owl:thumbnail	• http://commons.wikimedia.org/wiki/Special:FilePath/Light_dispersion_of_a_mercury-vapor_lamp_with_a_flint_glass_prism_IPNr°0125.jpg?width=300
dbpedia-owl:wikiPageExternalLink	• http://optics.byu.edu/textbook.aspx • http://www.cvimellesgriot.com/products/Documents/TechnicalGuide/fundamental-Optics.pdf • http://www.iop.org/publications/iop/index.html • http://www.iop.org/publications/iop/2009/page_38205.html • http://www.lightandmatter.com/area1book5.html • http://www.epic-assoc.com • http://www.myeos.org • http://www.optics2001.com • http://www.osa.org • http://www.osaindia.org • http://www.spie.org
dbpedia-owl:wikiPageID	• 22483 (xsd:integer)
dbpedia-owl:wikiPageRevisionID	• 609920324 (xsd:integer)
dbprop:hasPhotoCollection	• http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/Optics
dcterms:subject	• category:Applied_and_interdisciplinary_physics • category:Electromagnetic_radiation • category:Natural_philosophy • category:Optics
rdfs:comment	• Optics is the branch of physics which involves the behaviour and properties of light, including its interactions with matter and the construction of instruments that use or detect it. Optics usually describes the behaviour of visible, ultraviolet, and infrared light.
rdfs:label	• Optics
owl:sameAs	• freebase:Optics • http://nl.dbpedia.org/resource/Optica • http://fr.dbpedia.org/resource/Optique • http://de.dbpedia.org/resource/Optik • http://wikidata.dbpedia.org/resource/Q14620 • http://cs.dbpedia.org/resource/Optika • http://el.dbpedia.org/resource/Οπτική • http://es.dbpedia.org/resource/Óptica • http://eu.dbpedia.org/resource/Optika • http://id.dbpedia.org/resource/Optika • http://it.dbpedia.org/resource/Ottica • http://ja.dbpedia.org/resource/光学 • http://ko.dbpedia.org/resource/광학 • http://pl.dbpedia.org/resource/Optyka • http://pt.dbpedia.org/resource/Óptica • http://wikidata.org/entity/Q14620 • http://sw.cyc.com/concept/Mx4rv0dym5wpEbGdrcN5Y29ycA • http://af.dbpedia.org/resource/Optika • http://ar.dbpedia.org/resource/بصريات • http://az.dbpedia.org/resource/Optika

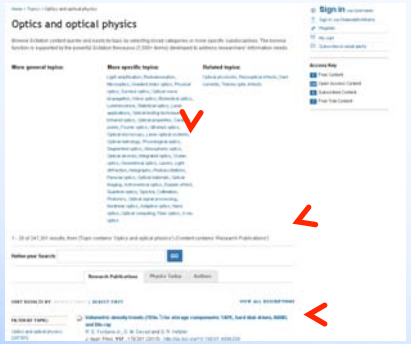
Access Innovations Inc.®
maker of Data Harmony taxonomy software

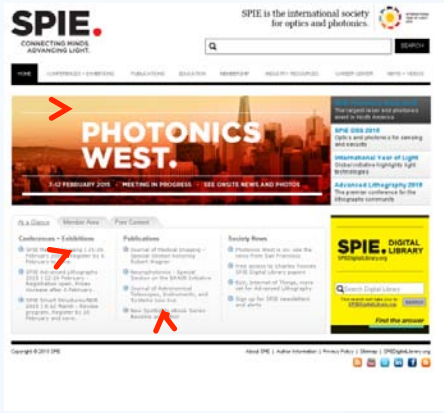
DataHarmony
from Access Innovations, Inc.

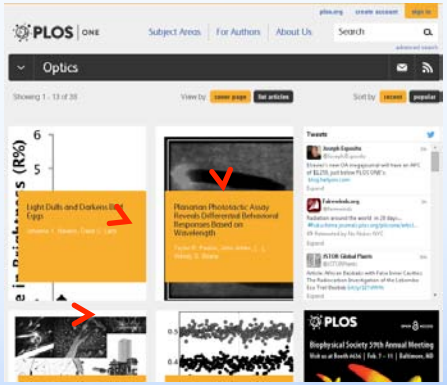
nesaurus


File Edit View Help
nomenclature
nondestructive readout
nonlinear media
nonlinear systems
nonprofit organisations
nozzles
nuclear electronics
nuclear facility decommissioning
nuclear facility regulation
nuclear materials packaging
nuclear screening
nucleation
nucleus
OBIC
occultations
oceanography
office environment
offshore installations
opalescence
open loop systems
open systems
operations research
optical bistability
optical control
optical elements
optical lattices
optical projectors
optical properties
optical resolving power
optical workshop techniques
optics
optimal systems
optimisation
Term: optics
rules for this term
Broader Term
Narrower Term
adaptive optics
aspherical optics
atmospheric optics
bio-optics
Fourier transform optics
geometrical optics
gradient index optics
integrated optics
micro-optics
nanophotonics
nonlinear optics
physical optics
quantum optics
Status Candidate Accepted
Related Term
aberrations
holography
lamps
laser beams
lasers
light
optical communication
Scope Note
External Link
External Link
http://dbpedia.org/page/Optics
Close
About: Optics
An Entity of Type : Thing, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org
Property
Value
dbpedia-owl:abstract
Editorial Note
Definition
DataHarmony
from Access Innovations, Inc.

One Concept, Unlinked



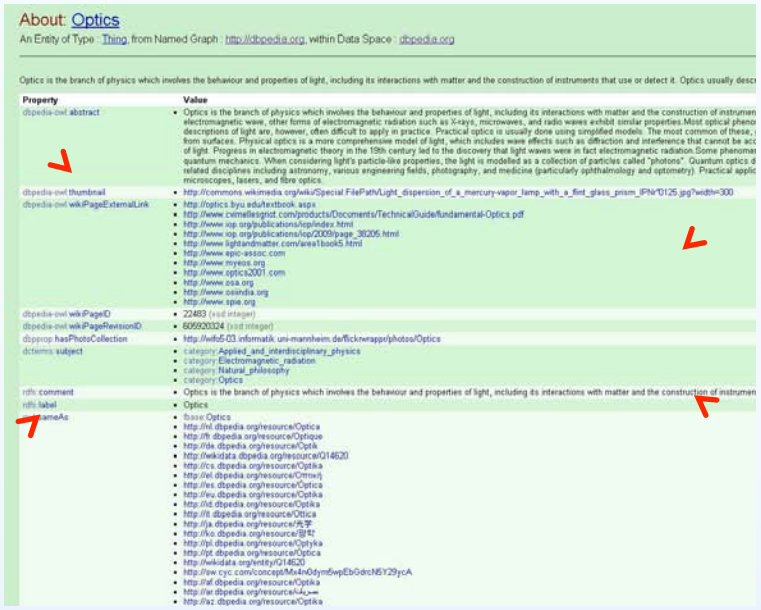


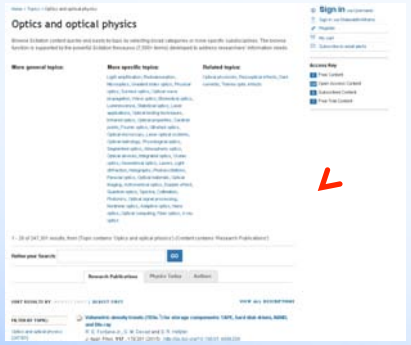


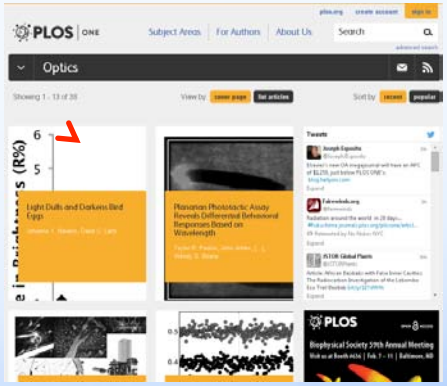
Access Innovations Inc.®
maker of Data Harmony taxonomy software

DataHarmony
from Access Innovations, Inc.

Point to a Persistent External Source







Access Innovations Inc.®
maker of Data Harmony taxonomy software

DataHarmony
from Access Innovations, Inc.

Automating the Linking

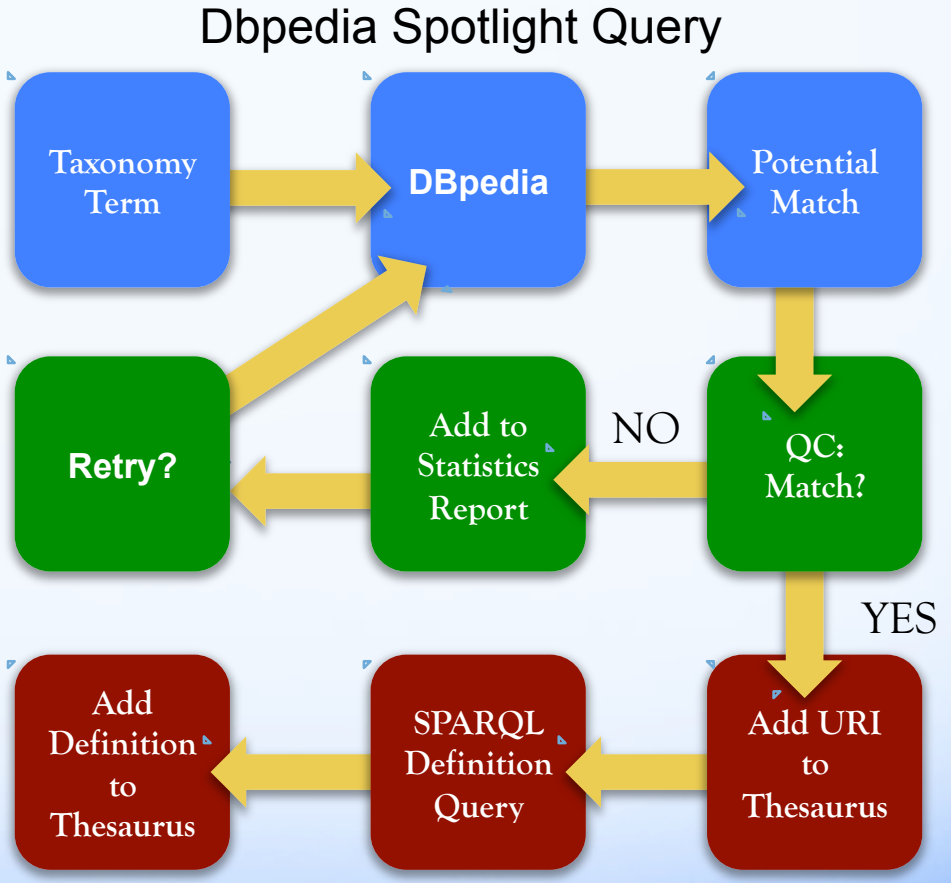
- Not every concept will have a match
- Or a resource page
- Semantic functionality –
  - Lots of synonyms will help
  - Proximity and other rules
- Create new resources or landing pages

Phrasing of Concepts Will Vary

- Exact concept match
  - Add the URI to a field in the thesaurus
- Different phrasing
  - Research funding/“Funding of Science”
  - SILK http://personal.sirma.bg/vladimir/misc/silk-book.pdf
- False matches
  - Ecosystem engineer vs. *Ecosystem engineering*

Linking Workflow

- Link content to external databank
- Make Potential URI matches
- QC for the thesaurus domain
- Matched URIs enrich the content



Access Innovations Inc.
maker of Data Harmony taxonomy software
DataHarmony
from Access Innovations, Inc.

Linking to External Datasets

- Link Thesaurus Preferred Terms
  - Resource describing the thesaurus concept
  - SKOS parlance, is “the same as”
- Identify DBpedia pages for each term
- Identify other sources
- Backfill knowledge gaps
- Concept exists
  - No content pages yet available

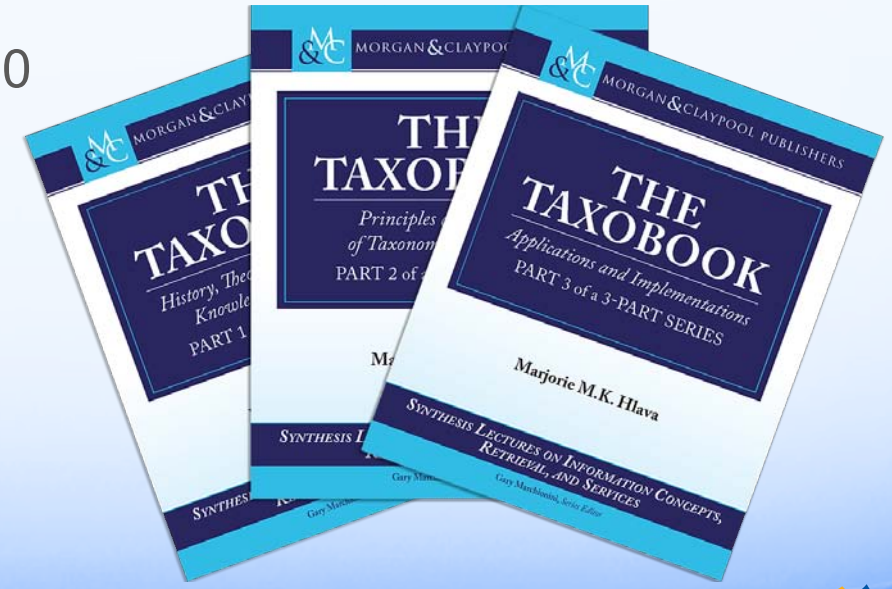
Application of a Smart Thesaurus

- Enriched inline tags
  - Launch a search
  - Redirect to topic page
- Links to other data sources
- External content on a page
- Topic Recommendations
- Images
- Recognition from Google, Google Scholar, etc.

Questions?

Thanks!

Marjorie M K Hlava, President
Access Innovations, Inc.
4725 Indian School Rd, Ste. 100
Albuquerque, NM 87110 USA
+1-505-998-0800
www.accessinn.com
www.dataharmony.com
Email: info@accessinn.com



Access Innovations Inc.®
maker of Data Harmony taxonomy software

DataHarmony
from Access Innovations, Inc.