The Role of Descriptive Metadata & Controlled Vocabularies in Improving Semantic Linkages

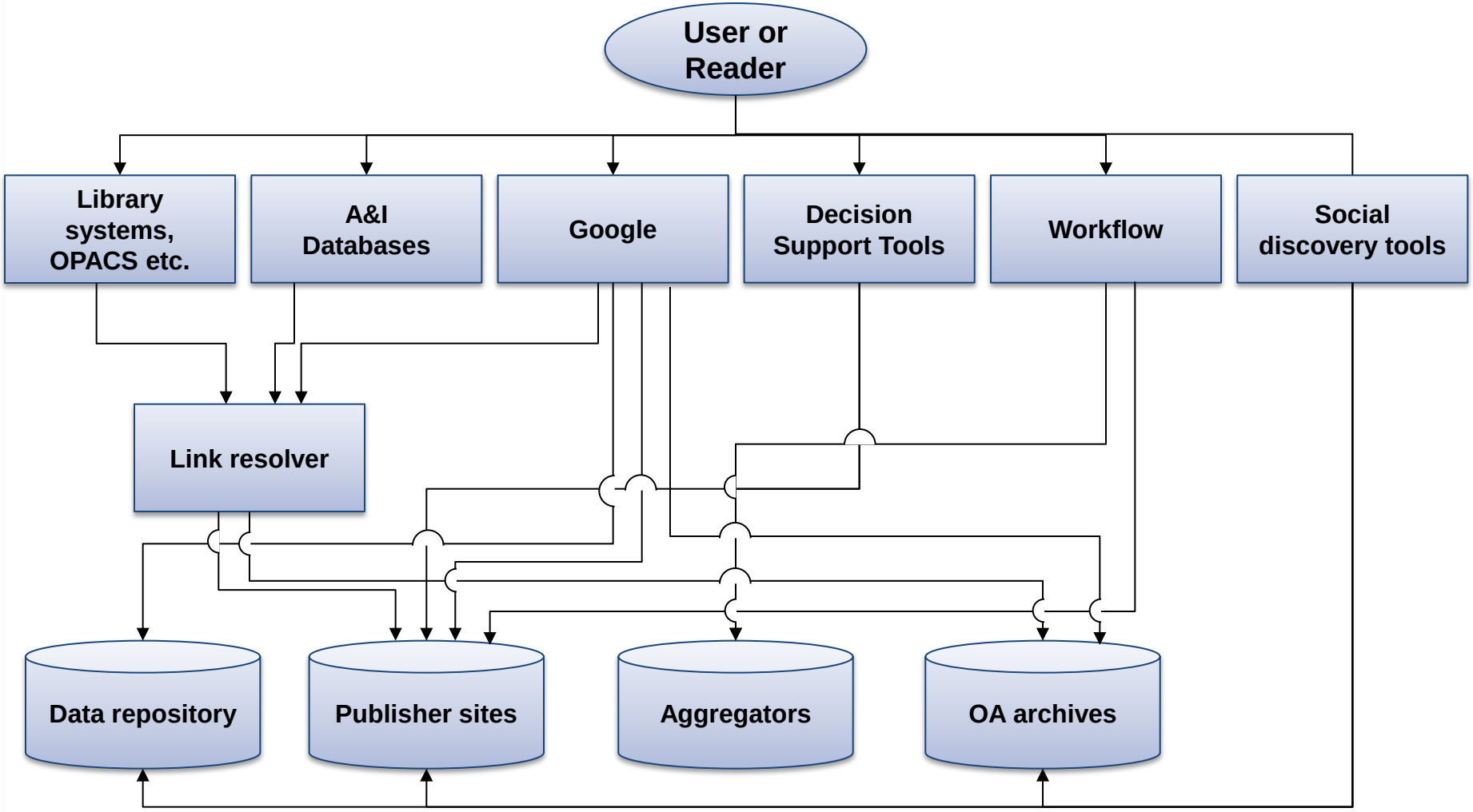
STM e-Production Seminar 2014

Narendra Venkatramani (Venky)
Senior Vice President, Sales & Marketing

SCOPE
eknowledge center
A Quatrro Global Services Company
mindsatwork

© Copyright Scope e-Knowledge Center 2014

Delivery Channels Have Become Complex



Enriched Metadata has become key for discoverability

© Copyright Scope e-Knowledge Center 2014

Types of Metadata

Bibliographic Metadata

- Title
- Author
- Other Identifiers (DOI, ISBN, Volume, etc.)

Descriptive Metadata

Content Structuring

- Subject Classification
- Subject Keywords
- Abstracts / Summaries
- Taxonomy / Thesaurus

Machine-aided human processing

Semantic Metadata

Content Intelligence

- Concepts
- Entities
- Semantic Linkages
- RDF & Triples

Knowledge modelling, intelligent agents

Increasing User Experience And Better Discoverability

© Copyright Scope e-Knowledge Center 2014

Role of Metadata in Discoverability

- **Helps enhance discoverability from document level to content level**
- **Helps in consistency in content discoverability through standardized indexing processes**
- **Increased monetization through deep discovery and multiple downloads of documents at more granular level e.g. chapters in books, tables and figures and other supplementary information**
- **Content powered with intelligence to automatically generate more meaningful relationships**

© Copyright Scope e-Knowledge Center 2014

Industry (NISO, Nielsen) Recommendations for Enriched Metadata

NISO Recommends Content Providers to provide enriched content (Abstracts & Keywords) to Discovery Service Providers for better discoverability

NISO RP-19-2014, Open Discovery Initiative

3.2.1.3 Enriched Content

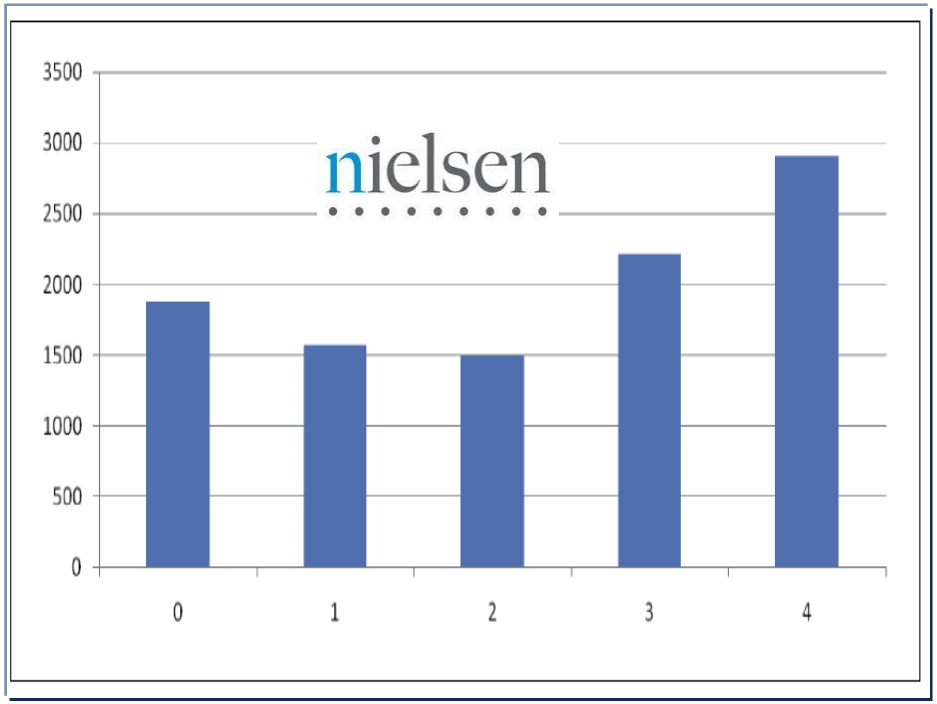
The elements in Table 3 may be provided by CPs to DSPs in total or in part for each item provided to a DSP for indexing. Inclusion of enriched content in indexes and as used for relevancy ranking greatly improves the discovery experience for users; it brings particular benefit to librarians and advanced researchers who are accustomed to controlled vocabularies. Examples are provided for each metadata element in Table 4.

Table 3: Enriched content to be provided by content providers

Field Name	Definitions
Indexing data	One or more keywords (from controlled or uncontrolled vocabularies) to describe the content of the item.
Full Text/Transcript	For text items, the entirety of the document. For audio or video content, a full transcript of the spoken content of the material. May not be relevant for all indexed content.
Abstract/ Description	Either a text summary on the content or (for non-text materials) a description of the item.

Source : NISO Open Discovery Initiative

High impact on sales for the books with enhanced metadata elements (short description, long description, review and autobiography) than books with less metadata elements



This graph shows the average sales per ISBN for records holding zero to all four enhanced metadata elements – short description, long description, review and autobiography.

Source : White Paper by Nielsen – The Link Between Metadata and Sales

© Copyright Scope e-Knowledge Center 2014

Definitions – Descriptive Metadata, Controlled Vocabularies (CV) & Semantic Linkages

Descriptive Metadata	▪ Describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords
Controlled Vocabularies (CV)	▪ Describes the subject and topics of documents using standardized terminologies
Semantic Linkages	▪ Relating documents using conceptual meaning rather than keywords

© Copyright Scope e-Knowledge Center 2014

Metadata Extraction at E-Production Stage (Automatic Extraction ???)

SCOPE eknowledge center
A Quatrro Global Services Company
minds@work

TOC as Abstract

Renewable Energy
Landolt-Börnstein - Group VIII Advanced Materials and Technologies Volume 3C, 2006, pp 1-10

1 Survey of renewable energy

K. Heinloth

Abstract

This document is part of Subvolume C 'Renewable Energy' of Volume 3 'Energy Technologies' of Landolt-Börnstein Group VIII 'Advanced Materials and Technologies'.

It contains:

1 Survey of renewable energy

1.1 Introduction: Renewable energy in the past, at present and in the future

1.2 Worldwide demand on energy and potential of renewable energy

1.3 Technologies to convert renewable primary energy

1.4 Possible utilization of the different kinds of renewable energy

1.4.1 Hydropower to produce electricity

1.4.2 Wind energy

1.4.3 Sun light

1.4.4 Biomass

1.4.5 Heat from earth interior

1.4.6 Heat from water, soil and air

1.4.7 Survey of main obstacles to increase the use of renewable energy by large

1.5 Synergy effects of extensive use of renewable energy

Generic Keywords (extracted from full text) without weightage for Relevance

Keywords

renewable
energy
electricity
power
energy technology
wind
hydropower
power plant
plant
HPP
fuel
water
photovoltaic
PV
solar
sun
biomass
geothermal
soil
economic aspects
hydro
river
turbine
Francis
Kaplan
Pelton
cavitation
head
thermal
tidal
TPP
Three Gorge Project
sluice
weir

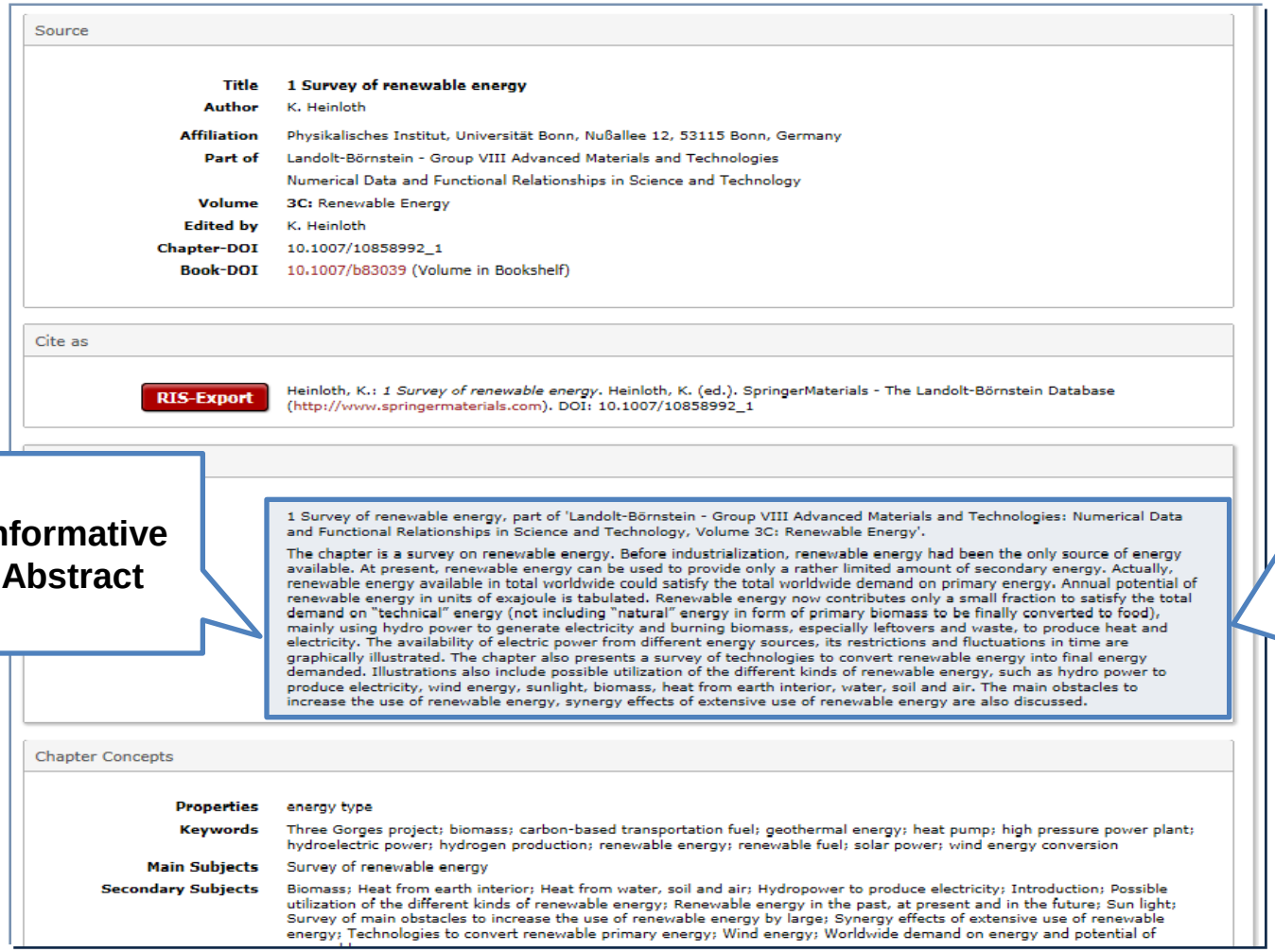
reservoir
pumpwater
electric
heat
hydrocarbon fuel
hydrogen
heat pump
generator
pump
discharge
dam
environment
bio
gas
diesel
biodiesel
ethanol
methanol
conversion
grid
radiation
silicon
conversion efficiency
collector
storage
residues
crop
combustion
bioenergy
gasification
furnace
electrolysis
electrolyze
photochemical
cooling
compression
carbon
dioxide
emission

emission
oil
reforming
synthesis
fermentation
ferment
pyrolysis
photo
alcohol
resources
earth
advection
heat capacity
thermal conductivity
temperature
bore hole
exchanger
hydrothermal
pollution
regenerative
regeneration
wood
reservoir
lining
cavern
reflector
receiver
efficiency
bio-gas
synfuel

Metadata extracted at production stage may not help in better discoverability

© Copyright Scope e-Knowledge Center 2014

Value Addition - Abstracting & Indexing



Source

Title	**1 Survey of renewable energy**
Author	K. Heinloth
Affiliation	Physikalisches Institut, Universität Bonn, Nußallee 12, 53115 Bonn, Germany
Part of	Landolt-Börnstein - Group VIII Advanced Materials and Technologies Numerical Data and Functional Relationships in Science and Technology
Volume	**3C:** Renewable Energy
Edited by	K. Heinloth
Chapter-DOI	10.1007/10858992_1
Book-DOI	10.1007/b83039 (Volume in Bookshelf)

Cite as

RIS-Export

Heinloth, K.: *1 Survey of renewable energy*. Heinloth, K. (ed.). SpringerMaterials - The Landolt-Börnstein Database (http://www.springermaterials.com). DOI: 10.1007/10858992_1

1 Survey of renewable energy, part of 'Landolt-Börnstein - Group VIII Advanced Materials and Technologies: Numerical Data and Functional Relationships in Science and Technology, Volume 3C: Renewable Energy'.

The chapter is a survey on renewable energy. Before industrialization, renewable energy had been the only source of energy available. At present, renewable energy can be used to provide only a rather limited amount of secondary energy. Actually, renewable energy available in total worldwide could satisfy the total worldwide demand on primary energy. Annual potential of renewable energy in units of exajoule is tabulated. Renewable energy now contributes only a small fraction to satisfy the total demand on "technical" energy (not including "natural" energy in form of primary biomass to be finally converted to food), mainly using hydro power to generate electricity and burning biomass, especially leftovers and waste, to produce heat and electricity. The availability of electric power from different energy sources, its restrictions and fluctuations in time are graphically illustrated. The chapter also presents a survey of technologies to convert renewable energy into final energy demanded. Illustrations also include possible utilization of the different kinds of renewable energy, such as hydro power to produce electricity, wind energy, sunlight, biomass, heat from earth interior, water, soil and air. The main obstacles to increase the use of renewable energy, synergy effects of extensive use of renewable energy are also discussed.

Chapter Concepts

Properties	energy type
Keywords	Three Gorges project; biomass; carbon-based transportation fuel; geothermal energy; heat pump; high pressure power plant; hydroelectric power; hydrogen production; renewable energy; renewable fuel; solar power; wind energy conversion
Main Subjects	Survey of renewable energy
Secondary Subjects	Biomass; Heat from earth interior; Heat from water, soil and air; Hydropower to produce electricity; Introduction; Possible utilization of the different kinds of renewable energy; Renewable energy in the past, at present and in the future; Sun light; Survey of main obstacles to increase the use of renewable energy by large; Synergy effects of extensive use of renewable energy; Technologies to convert renewable primary energy; Wind energy; Worldwide demand on energy and potential of

Informative Abstract

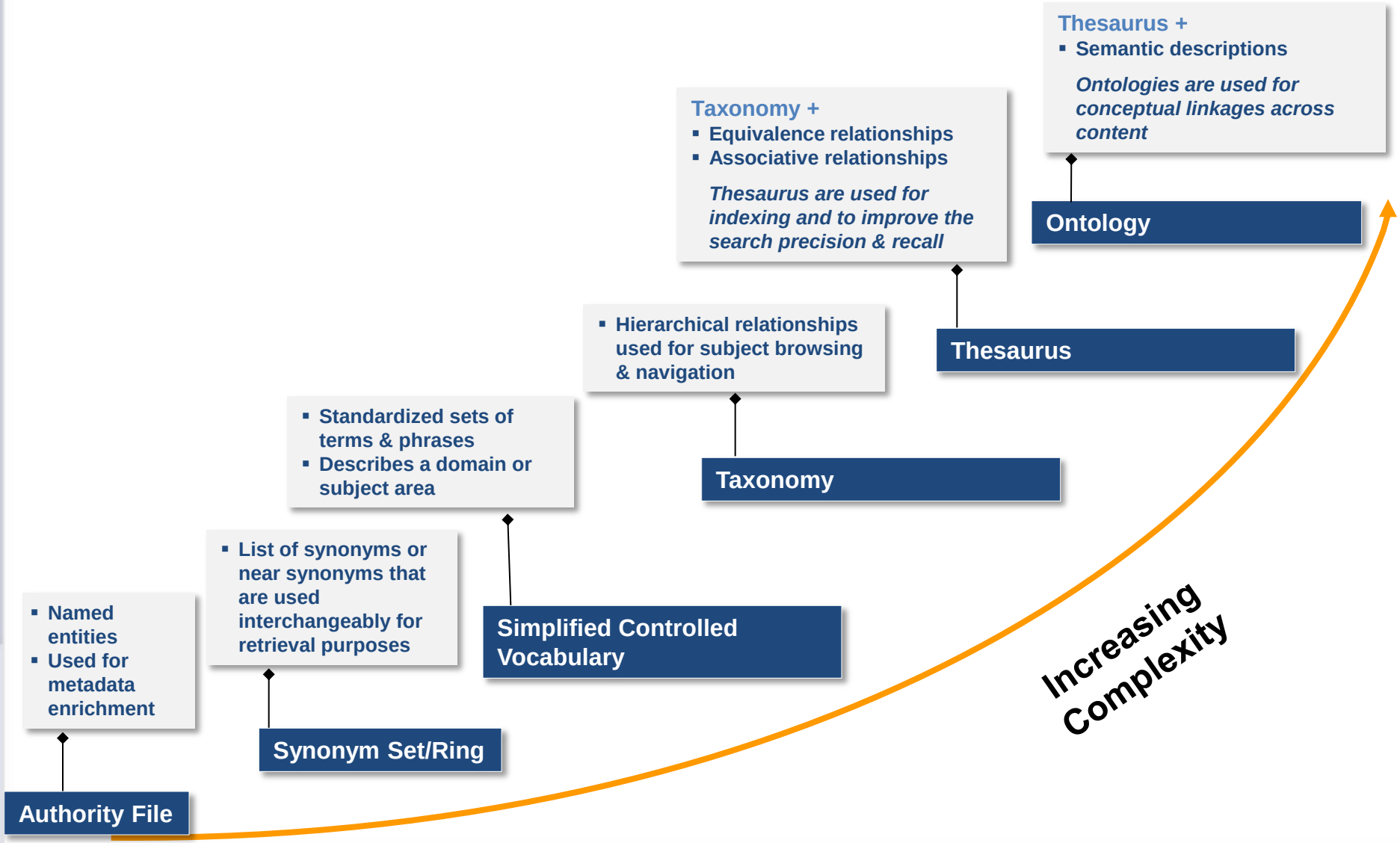
CV Terms:

- BIOMASS
- Electricity production
- Electrification
- Renewable Energy Sources
  Supply & Demand

Metadata enrichment through informative abstracts and CV indexing helps to enhance the relevance in discoverability

© Copyright Scope e-Knowledge Center 2014

The Controlled Vocabulary Value Chain

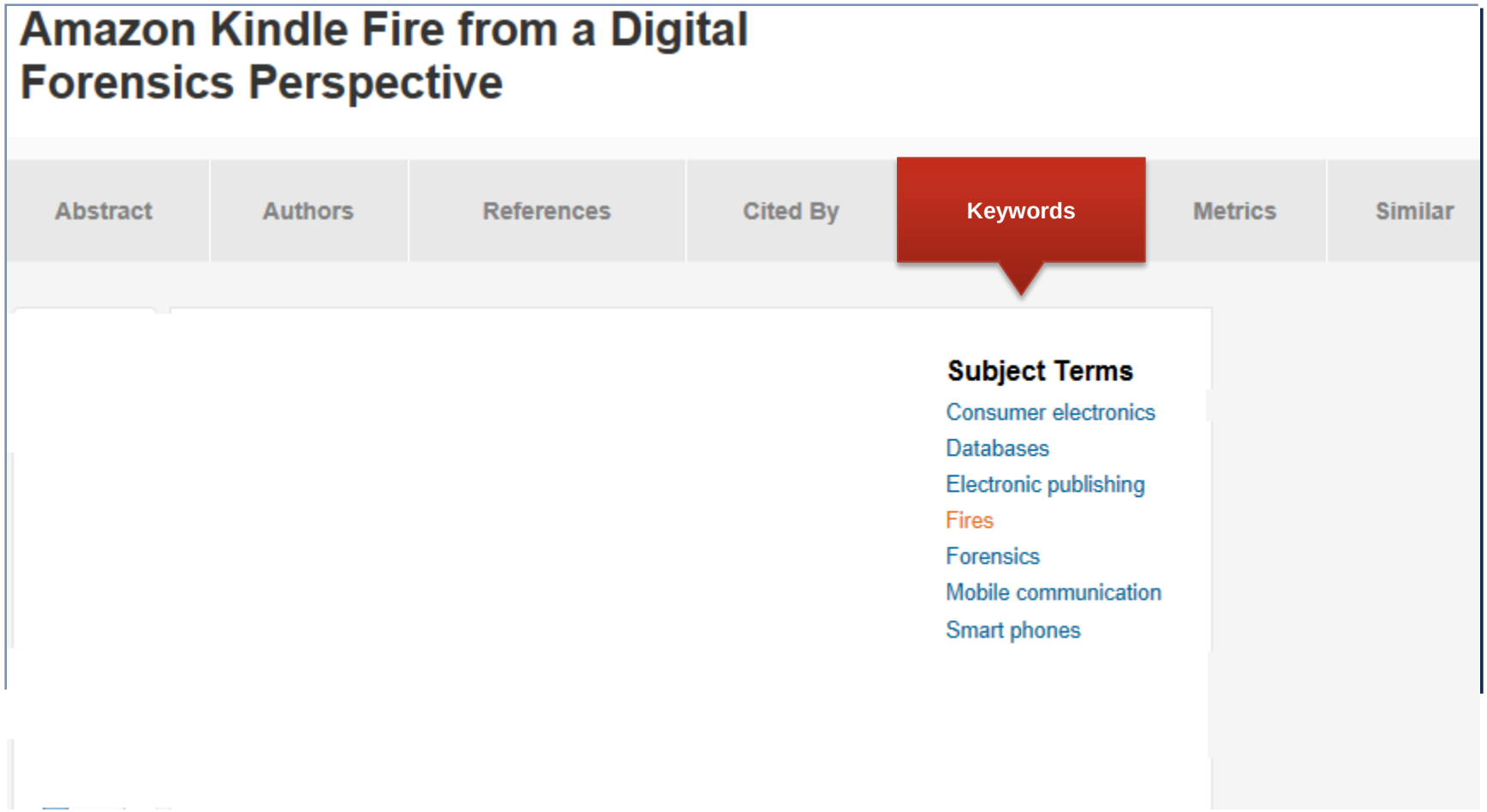


© Copyright Scope e-Knowledge Center 2014

Controlled Vocabularies & Semantic Linkages – Keyword Extraction as the Starting Point

© Copyright Scope e-Knowledge Center 2014

Controlled Vocabularies & Semantic Linkages – Automatic Extraction of Keywords Mapped to CV



© Copyright Scope e-Knowledge Center 2014

Controlled Vocabularies & Semantic Linkages – Looking for Context

Amazon Kindle Fire from a Digital Forensics Perspective

Abstract | Authors | References | Cited By | **Keywords** | Metrics | Similar

Forest Fires Regional Distribution and Numerical Simulation Based on Universal Kriging Algorithm
Guang Lu ; Wei Xue
Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on
Volume: 5
DOI: 10.1109/FSKD.2009.691
Publication Year: 2009 . Page(s): 110 - 113

Quick Abstract | PDF (337 KB) | HTML

Estimation of Biomass Burned Areas Using Multiple-Satellite-Observed Active Fires
Xiaoyang Zhang ; Kondragunta, S. ; Quayle, B.
Geoscience and Remote Sensing, IEEE Transactions on
Volume: 49 , Issue: 11 , Part: 2
DOI: 10.1109/TGRS.2011.2149535
Publication Year: 2011 , Page(s): 4469 - 4482
Cited by: Papers (2)

Quick Abstract | PDF (1709 KB) | HTML

Subject Terms

- Consumer electronics
- Databases
- Electronic publishing
- Fires
- Forensics
- Mobile communication
- Smart phones

The term "Fires" in CV

UF	Flames
BT	Hazards
RT	Accidents, Emergency services, Flammability, Hazardous areas Safety, Smoke detector

The Term "Fires" is automatically extracted and mapped with a CV in a wrong context of 'Forest Fires', while the article is on a electronic device called Amazon Kindle Fire. This results in irrelevant retrieval of related articles. Proper use of CV with display of broader and narrower terms helps eliminate such out-of-context terms (Fires) and related content (Forest Fires).

© Copyright Scope e-Knowledge Center 2014

Controlled Vocabularies & Semantic Linkages – How Abstracts Improves Contextual Relevance

Title

Amazon Kindle Fire from a Digital Forensics Perspective

Abstract

With the move toward **mobile computing** being the trend of this technology era it is clear that our way of life and how we deal with objects in it is changing. This swift shift from large desktop computers to inexpensive, low power applications that are easily carried in our pockets or placed next to a cup of coffee on the living room table clearly changed the way we interact with media and contact friends, colleagues and family members. This also created advancement in the field of **digital forensics** as with every device coming to the market, studies have been conducted to investigate the possible evidence that can be found on them. As we realize that with the comfort these devices do provide as a result of their mobility they are also providing a wealth of information about the users themselves for the same reason, hence they are really valuable **source of evidence** in an investigation. In this paper we will discuss one of these **mobile devices** which is **Amazon kindle Fire.** Being a new player in the **mobile computing** sector there haven't been enough studies of it in the field of **digital forensics** regarding it. In this paper we will discuss an **imaging** process to acquire the **data** from the device then we will provide an analysis of these data and their possible **sources of evidence.**

Abstract helps to identify the most relevant keywords on the topics discussed in the document

Keywords for Indexing

Controlled Indexing from CV

- Digital forensics
- Mobile computing
- Mobile device
- Databases

Free Indexing (From Author or Indexer)

- Amazon
- Kindle Fire
- Source of evidence
- Imaging

Semantic Fingerprint

Keywords can also be conceptually related to provide a semantic fingerprint of the document and retrieve more relevant related documents

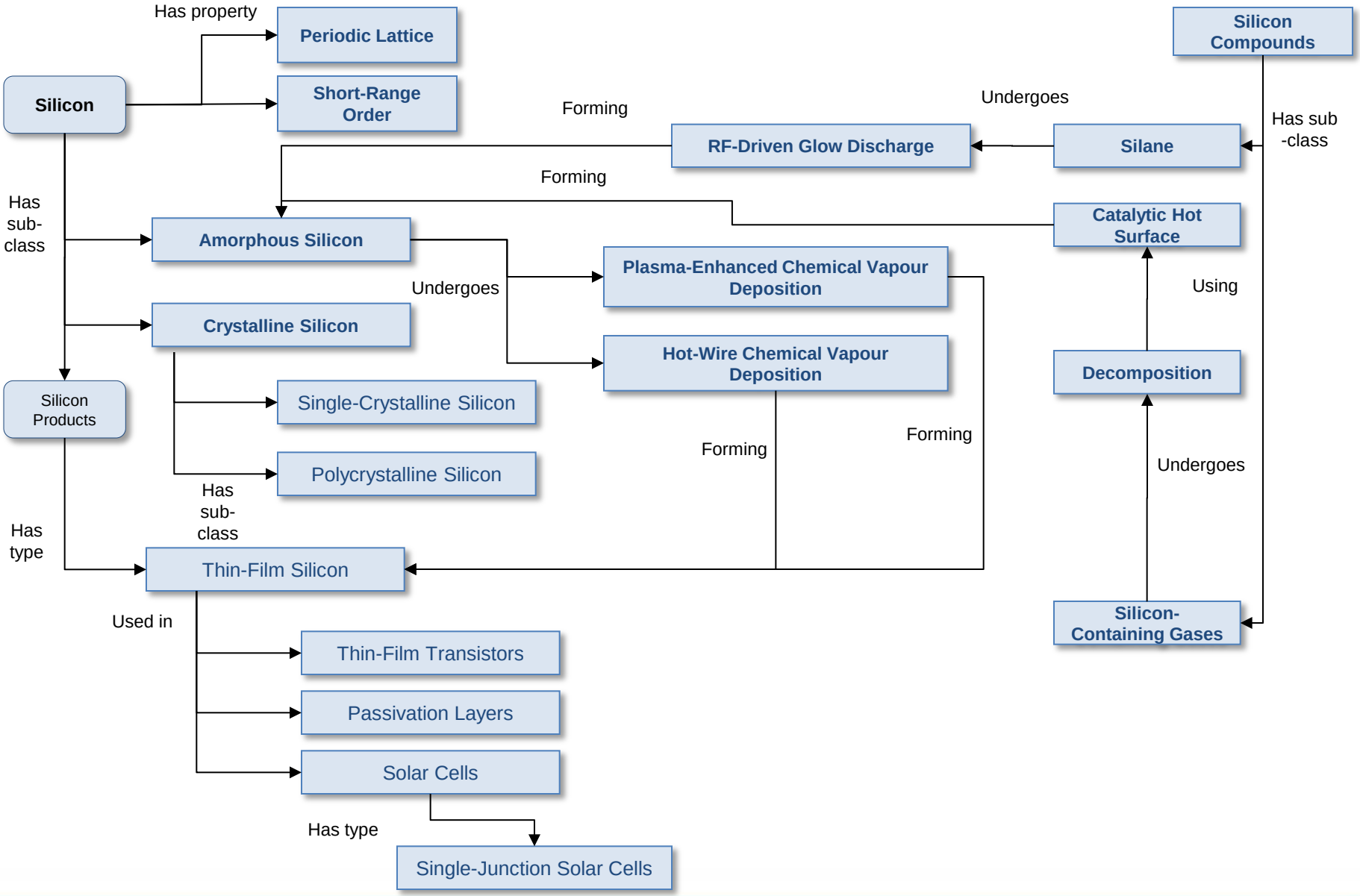
Kindle fire manufactured by Amazon is a mobile device used for mobile computing helps in generating databases extracted through an imaging process as source of evidence in digital forensics

Subject Terms for topical browsing

- Consumer electronics
- Mobile communication
- Forensics (Digital)

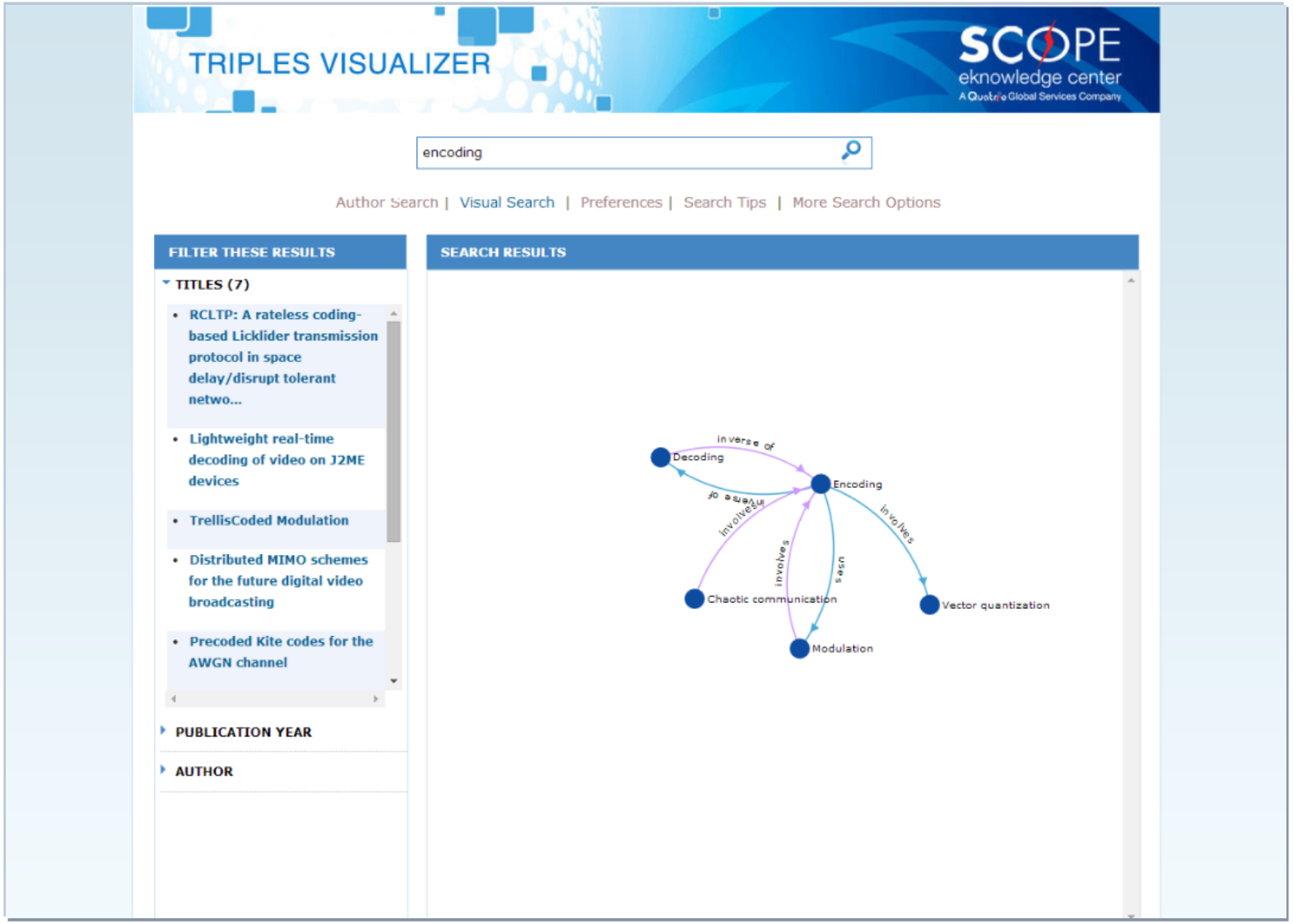
© Copyright Scope e-Knowledge Center 2014

Semantic Linkages – The Final Sprint to Knowledge Discovery



© Copyright Scope e-Knowledge Center 2014

Visualization of Semantic Linkages- Better User Experience



© Copyright Scope e-Knowledge Center 2014

How Metadata is Enriched for Different Users (for the Same Article)

Title	**Survey of renewable energy**
Author	**K. Heinloth**
Generic & Assorted keywords	Renewable, energy, electricity, power, energy technology, wind, hydropower, power plant, plant, HPP, fuel, water, photovoltaic, PV, solar, sun, biomass, geothermal, soil, economic aspects, hydro, river, turbine, Francis, Kaplan, Pelton, cavitation, head, thermal, tidal, TPP, Three Gorge Project, sluice, weir

Specific & Conceptual keywords

(Specialist Audience)

Properties	Renewable Energy types
Keywords	Three Gorges project; biomass; carbon-based transportation fuel; geothermal energy; heat pump; high pressure power plant; hydroelectric power; hydrogen production; renewable energy; renewable fuel; solar power; wind energy conversion

© Copyright Scope e-Knowledge Center 2014

About Scope

- ***Scope e-Knowledge Center –*** award winning global KPO provider of **Content Enrichment and Data Services**
- Over 600 Knowledge Professionals (Engineers, Scientists, Physicians, Scholars, Management Graduates, Librarians, Financial & Data Analysts)
- In-house IT team with an excellent track record in developing innovative techno-human platforms for the Publishing industry
- **prOdigi™ Lab** for conceptualization, fostering innovation and creating products/platforms aligned with customer demand

ISO 9001: 2008 (QMS) Since 2004

ISO 27001: 2005 (ISMS) Since 2005

Core Capabilities & Solutions

Content Enrichment and Discovery

- Medical Content Services
- Abstraction and Indexing Services
- Custom Taxonomy, Ontology and Classification Services
- Semantic Enrichment
- Usage Stats Analysis and Social Media

Data Services

- Entity Extraction & Normalization
- Authority Files Creation
- Data Acquisition ,Standardization, Normalization & Maintenance
- Business Research & Editorial Services
- Big Data Solutions

Deep Subject Matter Expertise

IT-Enabled Platforms

Multi-Lingual Capabilities

Technology-enabled knowledge partner to global enterprises

© Copyright Scope e-Knowledge Center 2014

Scope's Techno-human Platforms

SCOPE eknowledge center
A Quatrro Global Services Company
minds@work

ConSCIse™
A Unique Content Abstraction Solution

ConSCIse™, a unique content abstraction solution that delivers SEO Keyword-rich abstracts to enable greater discoverability

TranSCIse™
Non-English Content…Abstracted

TranSCIse™, a technology-enabled hybrid abstracting and indexing solution for enhancing the discoverability of non-English literature in search engines and A&I services

AuthEntik™
Author Data…Enhanced

AuthEntik™, an author data management solution capable of parsing, standardizing, normalizing and disambiguating a high volume of author information from original documents

InDEXr™
A Content Indexing Solution

InDEXr™, a platform-enabled indexing solution, offers a unique opportunity to enhance discoverability of content by providing highly-relevant indexing of key concepts / entities for a wide range of textual and non-textual documents

prodigi
Platforms ● Non Linearity ● Scale

Techno-human Model

- Service as a Product
- Scalable & Robust
- Non-linear Business Model
- Faster Go-To-Market
- Enhance Market Share

procuRx™
Spend Data.. Analyzed

procuRx™, a solution designed to provide and enhance the visibility of an organization's spend

OrdEHRSet™
Hospital Information System…Enabled

OrdEHRset™ offers faster, accurate and high-quality order sets that enable a healthcare information system to quickly reach maximum functionality and usability in meeting clinicians' needs

diSCOver™
Your Content...More Discoverable

diSCOver™, packages robust content mining, semantic enrichment and controlled vocabulary building and maintenance capabilities to improve discoverability and searchability

SemantiCz™
Concepts Linkages… Made Smarter

SemantiCz™, a platform based solution to enhance knowledge discovery through linked data of concepts and relationships using Ontologies and RDF

mARCat™
Meta Data…Catalogued

mARCat™, a platform-based solution to create MARC records for input files with dynamic cataloging template based on MARC standards, which will be validated by subject matter experts (SMEs)

Products for Publishers Products for Non Publishers (enterprises, portals, digital libraries)

Techno-human service platforms to enhance discoverability, enrich content and interpret knowledge

© Copyright Scope e-Knowledge Center 2014

Thank You

www.scopeknowledge.com
www.knowledgespeak.com

Narendra Venkatramani (Venky)
Senior Vice President, Sales & Marketing
venky@scopeknowledge.com

© Copyright Scope e-Knowledge Center 2014