American Journal of Computer Science and Information Engineering
2014; 1(1): 1-5
Published online August 10, 2014 (http://www.aascit.org/journal/ajcsie)

Keywords
Information Science Model,
Information Retrieval,
Data Analytics,
Knowledge Discovery,
Knowledge Management

Received: May 10, 2014
Revised: August 07, 2014
Accepted: August 08, 2014

The information science matrix: A theoretical framework for guiding knowledge discovery, knowledge management, data analytics and information retrieval

Harvey Hyman

Harvey Hyman, Florida Polytechnic University, 4700 Research Way, Lakeland, Florida 33805

Email address
hhyman@flpoly.org

Citation
Harvey Hyman. The Information Science Matrix: A Theoretical Framework for Guiding Knowledge Discovery, Knowledge Management, Data Analytics and Information Retrieval. *American Journal of Computer Science and Information Engineering.* Vol. 1, No. 1, 2014, pp. 1-5.

Abstract
This paper proposes a matrix for an information science framework for identifying and evaluating goals and objectives in the information problem space. The framework offers a model to represent four constructs within information science: knowledge discovery, knowledge management, data analytics and information retrieval. As a conceptual framework, the model supports the goals of understanding, evaluating and combining the four identified constructs into an integrated information science matrix (ISM). The matrix is a four-quadrant representation for a visual mental model of how one might think about the complexity and variety that make up information attributes for assessing the competing interests and goals of information oriented problem identification, specification and definition. The matrix is offered as a tool for managers and researchers seeking a framework to clarify understanding of problem sets in the domains of data analytics and information retrieval, further classified as knowledge management and knowledge discovery. The goal of the matrix is to serve as a reference model for improving problem definition and goal prioritization in the pursuit of information oriented applications.

1. Introduction

How do we *think* about information problems? If the goal of information is to produce knowledge, and if we divide that goal into the constructs of knowledge discovery and knowledge management on one axis, and the constructs of data analytics and information retrieval on a second axis, we create a two by two matrix to model four quadrants describing the information science landscape.

If we express these domains in terms of their descriptions and spheres of influence, we can create a model and framework to guide our thinking for describing information needs to define specific problem sets. There are four information functions that emerge from the interaction of the constructs of knowledge discovery, knowledge management, data analytics, and information retrieval: Definition, Organization, Explanation and Acquisition.

The ISM Framework suggests a configuration for a methodical approach for the many questions that have been, and continue to be investigated, and yet, still remain to be resolved or cohesively settled in a unified structure within the information

science domain: What does it mean to be "data driven?" How do we define the newly emerging term "Big Data Analytics?" How do we distinguish between information data, and knowledge? What is the difference between data analytics and information retrieval? How does knowledge management relate to knowledge discovery? What are the core foundational concepts common to all information problems, and how do they change when applied to different information applications. The ISM considers these questions; its purpose is to assist in defining, recognizing and defining these questions and serve as a platform to organize, specify, and develop delineated approaches to support applications within the information problem space.

2. Motivation

As the amount of information continues to grow and the economic incentives for "mining" and "analysing" data continues to increase, researchers and professionals alike struggle to define the landscape of information science. Currently, we find many academic and industry researchers grappling with the question; "We have all this data, now what do we do with it?" Quite often the dilemma lies in the inability to precisely define the information problem. The proposed ISM Framework is offered as a method to address this issue by offering an "information-oriented" approach to problem specification.

3. Two Paradigms: Data Analytics versus Information Retrieval

The ISM recognizes that information oriented problems can be divided into two distinct paradigms. The first is the paradigm of analytics; the second is the paradigm of retrieval. This is not a radical notion. In fact, it is self-evident. When we consider the goals and objectives of contemporary emerging information-oriented projects and initiatives, they can be classified into analysis oriented, such as data analytics, business intelligence and enterprise applications, or retrieval oriented, such as searching and sorting of documents and other information resources.

The analytic paradigm is observed most overtly in the recently popularized term "Big Data Analytics" and is operationally defined as the analysis of collections that are too large or too complex for conventional methods and techniques [1], [2]. This definition leaves out two important complexities that must be properly defined in order to perform competent analysis.

The first complexity is *defining the types of objects* within the collection. We need to know the type of data objects as well as their characteristics in order to specify the optimal method of handling and processing the set. The second complexity is to *identify and understand* the strategic information goals and objectives for what the researcher or management professional is attempting to achieve. This can lead to disparate goals that call for equally disparate, and sometimes ad hoc, exception handling methods and techniques [3].

The retrieval paradigm can be most overtly observed in new initiatives such as electronic medical records (EMR), corporate email preservation, and legal information retrieval (legal-IR). A specific application within legal-IR is known as eDiscovery, which is the exchange of digital information in litigation cases, and is estimated by some analysts to be a $2 billion industry.

The prototypical application within the paradigm of retrieval is Information Retrieval (IR). As such, it represents a completely different construct from Data Analytics. Whereas analytics is "data driven," meaning we are searching for a specific answer, relationship or pattern [4], [5], IR is "document driven," meaning we are searching for the presence or absence of documents [6]. Whereas, the analytical query seeks an answer through an emergent pattern or correlated relationship in the data [7], [8], the retrieval query seeks to find documents containing information relating to a subject [6].

4. The Knowledge Construct: Discovery versus Management

The ISM Framework considers how the two paradigms of analytics and retrieval interact with the two constructs of Knowledge Discovery (KD) and Knowledge Management (KM). By depicting a model for the interaction of the knowledge constructs along both paradigms, the ISM provides a way of thinking about how each knowledge construct differs in its approach based on which paradigm is defining the specific information problem defined. For example, a KD problem will have completely different and orthogonal goals and objectives if it is undertaking a retrieval project versus undertaking an analytics initiative. Likewise, a KM application will also have a different approach depending on whether its goal is analysis oriented or retrieval oriented. To further explain this relationship, let us consider the distinction between KD and KM constructs.

KD is concerned with the revelation of new relationships, previously unknown [9]. This information need has historically been implemented using a "mining" approach, leveraging applied methods in statistics, databases, machine learning and predictive algorithms [10].

KM on the other hand, is largely defined by the strategic goals of how organizations can aggregate, classify and codify their information knowledge base [11], [12], [13]. This is largely defined as an operational problem because it impacts how people and organizations leverage institutional knowledge and diffuse it throughout the organization so that people can maximize their effectiveness. We see examples of this in "thinking about thinking" [13], knowledge as a resource [12], and "enterprise knowledge

management" [11].

The ISM Framework presents a model to support information problem specification. For example, when a data analysis project is miss-specified as a knowledge discovery problem, but, in actuality, is a knowledge management problem, mismatched solutions emerge from poor definition of the original problem space.

Recognizing and correctly defining the problem space by using a tool such as the proposed ISM Framework, can lead to better matched goals and objectives for a research project, and therefore, improve the choices going forward and increase the rate of success in knowledge produced and achieved [14].

Another shortfall addressed by the ISM Framework is miss-specification between KD and KM initiatives. Quite often we specify our project as a KD problem, only to realize later that the problem really called for a KM approach to support organization, clarification or knowledge transfer [15].

Seeking out new, emergent patterns or previously unknown relationships may be modelled as an *explanation* problem, whereas seeking to organize large volumes of existing data may be modelled as a *definition* problem [16].

An IR project can be specified as a KD or KM problem. Similar to analytics projects, the IR implementation is optimally derived from the end goal desired. Seeking to amplify information through known sources may be modelled as an *organization* problem, whereas seeking new or more information on a subject may be modelled as an acquisition problem [17].

5. The Information Science Matrix as a Framework

Figure 1 depicts the ISM Framework for the Information Science landscape. We identify the four constructs of Knowledge Discovery (KD), Knowledge Management (KM), Data Analytics, and Information Retrieval (IR) -- these constructs represent the definitional descriptions in contemporary information-oriented problems. As explained previously, we classify information problems into two paradigms of analytics and retrieval and two constructs of knowledge discovery and knowledge management.

We define the four quadrants depicted within the matrix, formed by the convergence of the constructs along the two paradigms. We use operational terms to describe functional attributes that express the influence each construct has upon the defined information need, based on the classification between paradigms.

We describe the convergence of analytics and knowledge management (KM) as primarily *definition* oriented. Information problems within this quadrant are implemented as large-scale KM projects designed to achieve knowledge transfer [18]. We see these types of efforts reflected in models such as the SECI theory of knowledge transfer in business firms [16], [19].

The ISM describes a problem of information *definition* as one that is influenced by the goals of knowledge management and also classified as analytic oriented.

We describe the convergence of retrieval and knowledge management (KM) as goal focused information *organization*. Examples of information *organization* can be found in codification efforts to convert the *tacit* (implicit knowledge) to the *explicit* [19], [20].

The ISM describes a problem defined as information *organization* as one that is influenced by the goals of knowledge management and also classified as retrieval oriented.

We describe the convergence of analytics and knowledge discovery (KD) as goal focused information *explanation*. Information explanation describes problems defined by a user's need to better understand a topic or subject. This problem is represented as a puzzle and not as a mystery [21]. It is a puzzle because the user is not seeking an increased amount of information, but instead requires a clarification of the information already present.

The ISM describes a problem defined as information *explanation* as one that is influenced by the goals of knowledge discovery and also classified as analytics oriented.

We describe the convergence of retrieval and KD as goal focused information *acquisition*. This case is best described as a mystery rather than a puzzle [21] – the opposite approach to analytics and KD. In this instance, the user *is* seeking to increase the amount of information available on a topic or subject with the goal of determining the boundaries and frontiers of the knowledge base. Common examples of this quadrant are implemented as information retrieval applications in: Legal-IR (eDiscovery and litigation documents as described previously); Medical-IR in the form of electronic medical records and electronic health records (EMR/EHR); Freedom of Information Act (FOIA) requests [17].

The ISM describes a problem defined as information acquisition as one that is influenced by the goals of knowledge discovery and also classified as retrieval oriented.

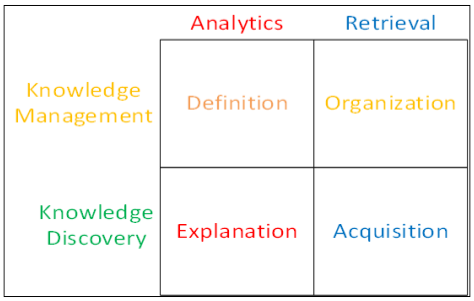

Figure 1. *The Information Science Matrix Framework*

6. A Structured Approach

The ISM Framework offers a holistic approach to information-oriented problems. It takes a three dimensional

view of information applications and their use. By classifying information problems into the paradigms of analytics and retrieval the ISM recognizes the unique requirements and constraints associated with the attributes that define the goals and objectives of an analytics initiative versus the goals and objectives of a retrieval project. By providing descriptive functions that explain the convergent and disparate nature of projects that are analytic versus retrieval oriented and knowledge management versus knowledge discovery in objective, the ISM offers a comprehensive model for improving information problem definition and supports an information-oriented approach to specification and resolution in applications and use.

The ISM Framework can assist managers and researchers to organize their schema for information projects by providing a model that offers a new way to think about the competing interests in an information problem. By balancing the dual paradigms of analytics and retrieval, and allowing for overlap and distinction between knowledge management and knowledge discovery, the ISM makes it possible to define and appreciate the contributions and limitations associated with the four identified functional quadrants that emerge in the information science domain.

The functional descriptions of information: definition, organization, explanation, and acquisition explain the operational needs associated with the interactions of each information construct with each of the paradigms.

The corresponding attributes within each information functional quadrant support a multi-faceted approach to information problems. The descriptive functions are designed to promote broad thinking about research questions and industry practices motivated by the emerging information needs in contemporary applications and use.

7. Information Oriented

The ISM Framework supports improved critical thinking and problem definition for information projects and applications. By thinking about problems as "information-oriented," we immediately frame our view of the problem space in a defined, unambiguous way. We state the problem as an "information problem" and the corresponding goals as "information goals."

The functional descriptions (definition, explanation, organization, acquisition), offer explanatory images that serve as metaphors for thinking about problems as *information* problems, and setting goals and objectives as *information* goals and objectives.

Examples of definition questions the ISM Framework can support for an improved information-oriented approach are: What is the suggested method of sorting, aggregating and reporting the information in this domain? How will the system handle context determinacy (mapping and definitions) for this problem set? What is our method for content management (organization) of this data? What is our relevancy model to define the information need in this instance?

8. Conclusion

We offer the Information Science Matrix (ISM) Framework for consideration in defining information problems and setting goals and objectives. The purpose of this paper is to stimulate additional thinking and debate about how we can better define and structure the Information Science landscape and produce improved problem definition, enhanced research, and substantial impacts on industry approaches to information classification and knowledge production.

The ISM Framework is a new model for thinking about how current and emerging information-oriented problems can be defined and specified by using a unified approach to the knowledge constructs and the information paradigms.

As information and data continue to expand, in terms of volume and availability, reference models such as the ISM Framework will become increasingly powerful in the realm of critical thinking and problem definition efforts.

Acknowledgements

The author would like to acknowledge the support of H2 & WF3 Research, LLC, and Omniscient Analytics Corporation for their contributions and efforts in supporting the on going work in this domain. The continued support of these two organizations makes this work possible.

References

[1] Ward JS and Barker A. Undefined by Data: A Survey of Big Data Definitions. Cornell University Library. arXiv:1309.5821v1 [cs.DB] 2013.

[2] Lany D. Application Delivery Strategies. Meta Group 2001.

[3] Hyman H S and Fridy W.Using Bag of Words and Standard Deviations to Represent Expected Structures for Document Retrieval: A Way of Thinking That Leads to Method Choice. NIST Special Publication: Text Retrieval Conference (TREC) Proceedings 2010.

[4] NIST Big Data Working Group 2013.

[5] Google Trends for Big Data 2013.

[6] Vanrijsbergen CJ. Information Retrieval.Butterworths 1979.

[7] Totty P. Unearth Hidden Treasues with Data Analytics. Credit Union Magazine 2014.

[8] Barton D and Court D. Making Advanced Analytics Work for You. Harvard Business Review 2012.

[9] Berndt DJ Bhat S Fisher JWHevner AR and Studnicki J. Data Analytics for Bioterrorism Surveillance.*Second Symposium on Intelligence and Security Informatics* ISI 2004: Proceedings, Nigorikawa K (ed).

[10] McBrien KA Kleinman KP Abrams AM Prosser LA. Use of Outcomes to Evaluated Surveillance Systems for Bioterrorist Attacks. BMC Medical Informatics and Decision Making 2010.

[11] Frawley WJ Piatetsky-Shapiro G and Matheus CJ. Knowledge Discovery in Databases: An Overview. AI Magazine 1992.

[12] Tiwana A. The knowledge management toolkit: practical techniques for building a knowledge management system. Prentice Hall 2000.

[13] AlaviM and Leidner D. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. Management Information Systems Quarterly 2001.

[14] Wiig KM. Knowledge management foundations: thinking about thinking: how people and organizations create, represent, and use knowledge:Vol. 1. Schema Press 1993.

[15] Buckeridge Dand Lombardo J. Disease Surveillance: A Public Health Informatics Approach. Wiley-Interscience 2007.

[16] Nonaka I. A Dynamic Theory of Organizational Knowledge Creation. Organization Science 1994.

[17] Hyman HS. Systems Acquisition, Integration and Implementation: For Engineers and IT Professionals. Sentia Publishing 2014.

[18] Hyman HS and Fridy W. An eHealth Process Model of Visualization and Exploration to Support Improved Patient Discharge Record Understanding and Medical Knowledge Enhancement. Clef Conference 2014.

[19] Nonaka I. Toyama R and Konno N. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. Long Range Planning 2000; Volume 33, Issue 1.

[20] Polyani. Personal Knowledge: Towards a Post Critical Philosophy. London: Routledge 1958.

[21] Gladwell M. What the Dog Saw 2009.