Bibliographic Classification in the Digital Age: Current Trends and Future Directions

Asim Ullah, Shah Khusro, and Irfan Ullah

ABSTRACT

Bibliographic classification is among the core activities of Library & Information Science that brings order and proper management to the holdings of a library. Compared to printed media, digital collections present numerous challenges regarding their preservation, curation, organization and resource discovery & access. Therefore, true native perspective is needed to be adopted for bibliographic classification in digital environments. In this research article, we have investigated and reported different approaches to bibliographic classification of digital collections. The article also contributes two evaluation frameworks that evaluate the existing classification schemes and systems. The article presents a bird's-eye view for researchers in reaching a generalized and holistic approach towards bibliographic classification research, where new research avenues have been identified.

INTRODUCTION

Classification is the primary instinct of human beings in arranging, understanding, and relating knowledge artifacts. Bibliographic classification provides a framework for arranging and organizing knowledge artifacts preserved in the form of books, magazines, newspapers and other holdings to explore new avenues of knowledge management. Today several classification schemes are in use ranging from conventional classification schemes including Library of Congress Classification (LCC), Dewey Decimal Classification (DDC), Colon Classification (CC), and Universal Decimal Classification (UDC) to classification for digital environments including Association for Computing Machinery (ACM) digital library[1], Institute of Electrical and Electronics Engineering (IEEE) digital library[2], and Online Computer Library Center (OCLC) cooperative catalogue[3].

Besides the difficulties that lie in devising a classification scheme (time-consuming and resource-consuming), it is required that either the existing schemes should be revised and extended or a new classification scheme should be devised, which could act as a common platform for representing knowledge artifacts belonging to different contexts. Such a classification scheme should also resolve the challenges in digital preservation and curation and support the precise

Asim Ullah (asimullah@upesh.edu.pk), **Shah Khusro** (khusro@upesh.edu.pk), and **Irfan Ullah** (cs.irfan@upesh.edu.pk) are researchers at the Department of Computer Science, University of Peshawar, Peshawar, Pakistan.

[1] http://dl.acm.org/
[2] http://ieeexplore.ieee.org/Xplore/home.jsp
[3] https://www.oclc.org/

And accurate search and retrieval of digital collections. The first step, in this connection, is to properly analyze and evaluate the existing bibliographic classification schemes and to dig out their strengths and limitations in classifying digital collections accurately and appropriately. Therefore, the objectives of this research article include:

- To investigate and evaluate the available approaches to bibliographic classification from the perspective of devising a classification scheme that can act as a common platform for classifying any type of digital collection.
- To devise evaluation frameworks that compares the available bibliographic classification schemes and approaches.
- To present issues, challenges, and research opportunities in state-of-the-art bibliographic classification research.

The rest of the paper is organized as: Section 2 presents the current trends in the classification of digital collections. Section 3 presents two evaluation frameworks for comparing and evaluating the existing solutions. Section 4 presents research challenges and opportunities in bibliographic classification research. Finally, Section 5 concludes our discussion. References are presented at the end of the paper.

Classifying Digital Collections – A Mixed Trend

The bibliographic classification has been the focus of several researchers to properly classify, catalogue, and describe digital collections. In this regard, two approaches have been adopted: the former supports the use of conventional classification schemes including CC, DDC, and LCC etc., in describing and classifying digital documents, while the latter recommends devising some new ways of classification such as ACM[4] computing classification. However, in most of the digital environments, a mixed trend has been observed, where along with new classification schemes, categorization is also used as a complementary solution. For example, ACM presents its own classification system as poly-hierarchical ontology in describing Computer Science literature and for using in Semantic Web applications. ACM has replaced its 2008 ACM classification system that serves as de-facto model for the classification of Computer Science literature by giving visual topic display along with searching services. It serves as semantic vocabulary for categorizing concepts and a foundation of computing disciplines ("The 2012 ACM Computing Classification System,"). Similarly, IEEE digital library categorizes its holdings into directories per its own rules of cataloguing and categorization. It categorizes articles and standards in to several subject areas and clusters documents through year of publication, author names, content type, affiliation, publication title, publisher, country of publication, alphabets, numerals and alphanumeric values[5]. The document collection can be navigated through collection names, number of documents, by topic and International Classification for Standards (ICS).

[4] http://dl.acm.org
[5] http://ieeexplore.ieee.org/browse/standards/ics/ieee/

doi:10.6017/ital.v36i3.8930

The DMOZ[6] directory is the largest human made directory of web pages. Since its inception in 1998, it categorizes 3,861,137 websites available in 90 languages into 1,031,719 categories and sub-categories by 91,928 editors and volunteers. In addition, it has its DMOZ RDF dumps available on Linked Open Data (LOD) cloud. According to the World Wide Web Consortium (W3C), LOD enables the data integration and reasoning at a large scale ("Linked data,"). It establishes links among data enabling machines and users to explore the web of data rather than the web of documents along with finding related data (Berners-Lee, 2006; Bizer, Heath, & Berners-Lee, 2009). However, it lacks in semantic search (meaningful search), which affects the precision and accuracy in exploring the required resources. Also, the categories, under which the websites are kept, are needed to be revised because there can be faceted and intra-hierarchical links among web pages. In addition, the content management needs to be upgraded for updating the directory with new entries and the way it reviews and categorizes websites (Boykin, 2016).

Institutional repositories use the mixed approach towards creating, collecting and managing metadata for printed and digital collections using several sources including conventional and digital. This mixed trend introduces challenges to the metadata managers (Chapman, Reynolds, & Shreeves, 2009). To deal with these challenges, the subject classification systems can be very beneficial in providing Web-oriented services including searching of contents through search patterns, browsing, and content filtering by subject area. However, at the same time, a cognitive overload rises for the authors and depositors of the institutional repository (Cliff, 2008) that needs further attention.

To handle the information overload in retrieving digital collections, several controlled methods have been proposed in the literature ranging from manual techniques (e.g., web directories) to automatic techniques including clustering and classification. Several classification schemes including sentiment and subject classification have been developed for classifying (and categorizing) web pages. Classification is used in focused crawling, searching and ranking results, and classifying queries. Clustering also classifies web resources but it is slightly different from classification, which is based on a rigid predefined taxonomy and rules for interpreting the meaning of classification order. On the other hand, clustering shows flexibility in classification (categorization) of web documents (Zhu, 2011). However, a mixed trend has been observed, where classification and categorization are intermingled to facilitate organization, description, exploration, and retrieval of digital collections.

Semantic Web brings meaningful connections between the web of data so that not only humans but machines can also understand the content of documents to retrieve the most intended and required documents. This way other related documents could also be easily connected and retrieved (Berners-Lee, 2006). To understand, describe, and relate concepts within documents, ontologies are used. Therefore, researchers have been working on bringing semantics through Semantic Web and related technologies to automatically classify digital collections. For example,

[6] http://www.dmoz.org

(Beghtol, 1986) argues that semantic axis makes syntactical classification structure more meaningful and provides the platform for developing relationships among knowledge artifacts through several warrants in classification systems. Similarly, classification ontology is used in automatic classification (Wijewickrema & Gamage, 2013) to minimize the ambiguity in vocabulary. To obtain a single subject for the input document, several weight functions including the term frequency-inverse document frequency (TF-IDF), and filtering methods are applied.

Semantic Web and LOD technologies have also been used in dealing with bibliographic data. For example, BibBase[7], a bibliographic data publishing and management tool (Xin, Hassanzadeh, Fritz, Sohrabi, & Miller, 2013) publishes bibliographic data on the user website according to LOD principles. However, these are limited because of the lack of interoperability among native languages while translating classification records from source language to the target language (Kwaśnik & Rubin, 2003).

The classification schemes are also being converted into ontologies. (Giunchiglia, Marchese, & Zaihrayeu, 2007) have applied reasoning capabilities of OWL ontologies to classification schemes. These ontologies are used as interfaces to human knowledge for machines whereas classification schemes are interfaces to knowledge for humans. However, there is limited support available for cross-disciplinary searching and accommodation for more views and interpretations of knowledge (Albrechtsen, 2000). The supervised and unsupervised machine learning techniques are used for automatic text classification. Supervised machine learning techniques use models including multinomial Naïve Bayes model, and Bernoulli model (Manning, Raghavan, & Schütze, 2008) for classification. Yelton (2011) applies probabilistic classification of important words (and therefore of documents) especially by considering Amazon's Statistically Improbable Phrases (SIPs)[8] and Google phrase search inside a book. For subject analysis, he mentions simplistic; content-based; and requirements-based methods in terms of understanding text classification and manipulation of books. The Wikipedia page structural hierarchy is exploited in automatically harvesting, classification, categorization, clustering, and metadata enrichment (Yelton, 2011).

Information Extraction (IE) is also applied in classifying books automatically. For example, (Betts, Milosavljevic, & Oberlander, 2007) use IE methods for automatic labeling of books using LCC classification. They used bag-of-words (BOW) model, bag-of-named-entity recognition (NER) model, generalizing named entities (GAZ) model in automatic text classification. To achieve better accuracy, they also combined the results of these models. However automatic classification may lead to limited search and retrieval because of the missing semantics associated with phrases or key words. To overcome this issue, a fundamental and practical theoretical model of classification is required (Jones, 1970).

[7] https://bibbase.org/
[8] http://www.amazon.com/gp/search-inside/sipshelp.html

doi:10.6017/ital.v36i3.8930

Table 1 categorizes the bibliographic classification approaches into three broader categories namely: theoretical approaches, practical approaches and approaches used in digital environments. Theoretically researchers have discussed different viewpoints for classification, whereas we get a different view when these schemes are applied for classification. Practically, the syntactic structure is valued by using faceted and enumerative techniques. In digital environments like the Web and digital libraries, strict boundaries of classification are often compromised by categorization.

Approaches to Classification	Techniques Used
Theoretical Approaches	1. Biasness (Mai, 2009) (Mai, 2010) 2. Subjectivity and objectivity (Hjørland, 2016) 3. Epistemological and Semiotic approaches (Hjørland, 2013) (Lee, 2012; Mai, 2011) (Tennis, 2008) 4. Empiricism, Rationalism, Historicism and Pragmatism (Hjørland, 2013) 5. Multidisciplinarity approach (Beghtol, 1998) 6. Scientific approaches (Hjørland, 2008) 7. Positivistic and pragmatic approaches (Dousa, 2009) (Mai, 2011) 8. Interdisciplinary and evidence based practice classification (Hjørland, 2016) 9. Social and cultural context (J.-E. Mai, 2004) 10. By tracking the universe of knowledge 11. Universal order (Smiraglia & Van den Heuvel, 2011) 12. Integrative levels in classification (Dousa, 2009) 13. Literary warrant (Rodriguez, 1984) 14. Education warrant (Hjørland, 2007) (Beghtol, 1986) 15. Semantic warrant (Beghtol, 1986) 16. Syntactic warrant (Beghtol, 1986) 17. Domain and users requirements (Mai, 2005) 18. Pluralism and human interpretations
Practical Approaches	1. Enumerative and Faceted (Batley, 2014). 2. General Purpose approach (Mai, 2003) and Special Purpose approach (Mancuso, 1994) e.g. classification schemes for general classes of knowledge areas or for a special class of knowledge area. 3. Syntactic axis (Beghtol, 1986) (Beghtol, 2001) 4. Semantic axis (Beghtol, 1986) (Beghtol, 2001)

Classification in Digital Environment	1. Document Similarity (Hamming distance and Euclidean geometric approaches) (Losee, 1993) 2. Fuzzy approach (Jacob, 2004) 3. Clustering (Nizamani, Memon, & Wiil, 2011) 4. Categorization (Koshman, 1993) 5. TF-IDF weighting (Dorji et al., 2011) 6. Unsupervised machine learning techniques (Joorabchi & Mahdi, 2011). (K-mean Clustering, hierarchical clustering) 7. Supervised machine learning techniques (Wang, 2009) (Multinomial Naïve BAYES, Bernoulli model, Support Vector Machine, Random Forest, K-NN technique) 8. Information Extraction methods (Gilchrist, 2015) 9. Probabilistic text and document classification (Maron, Kuhns, & Ray, 1959) 10. Ontologies (Campbell, 2002)

Table 1. Categorization of approaches towards bibliographic classification

Evaluating Classification Schemes & Approaches

In this Section, we present two evaluation frameworks to compare and evaluate the existing classification and categorization systems and well-known bibliographic classification ontologies. We have chosen CC, DDC, LCC, and Universal Decimal Classification (UDC) on the basis of their structural properties and wide usage both in conventional and digital libraries ("Subject classification schemes," 2015) ("Library of Congress Classification," 2014) ("About Universal Decimal Classification (UDC),") (Press, 2002) (Encyclopedia, 1 August 2014). Some of these properties include: citation and filling order; notations expressiveness; flexibility in classification principles, rules and notations; coverage of the knowledge areas; classification schedules and notations structure; notations brevity and simplicity; notations mnemonics; notations hospitality; schedules with updateable and comprehensive subjects order; and knowledge coverage (Batley, 2014). The UDC, LCC, and DDC are universal, multidisciplinary, and widely used systems (Koch & Day, 1997), whereas CC has the seminal and inspirational value for the faceted structure of the bibliographic classification. Therefore, the evaluation framework mainly targets these classification schemes as our natural choice for the evaluation and comparison. Similarly, we evaluate ACM[9], IEEE[10] and DMOZ[11] using the evaluation framework as these are the well-known and widely used document classification & categorization systems for the digital libraries. Table 2 presents 22 metrics used in the evaluation framework. These evaluation metrics are extracted

[9] http://www.acm.org/about/class
[10] http://www.ieee.org/about/today/at_a_glance.html?utm_source=mm_link&utm_campaign=iaa&utm_medium=ab&utm_term=at%20a%20glance
[11] https://www.dmoz.org/docs/en/about.html

doi:10.6017/ital.v36i3.8930

from the existing literature (Kaosar, 2008) (Painter, 1974) (Encyclopedia, 1 August 2014) (Buchanan, 1979) (Kaosar, 2008) (Painter, 1974) (Encyclopedia, 1 August 2014) (Koch et al., 1997) (Reiner, 2008) (Gnoli, Merli, Pavan, Bernuzzi, & Priano, 2008) (Francu, 2007) (Chan, Intner, & Weihs, 2016). These metrics include: (i) structural complexity; (ii) notational brevity; (iii) predefined structure; (iv) rules complexity; (v) theoretical laws; (vi) mnemonics; (vii) hospitality; (viii) search complexity; (ix) usability; (x) precision and accuracy; (xi) multilinguality; (xii) interoperability; (xiii) semantic search; (xiv) bias in subject representation; (xv) enumerative structure; (xvi) faceted structure; (xvii) faceted search; (xviii) consistency; (xix) LOD datasets; (xx) Linked Open Vocabularies (LOV) support; (xxi) platform; and (xxii) warrants of classification. These metrics, their need, and their use in ratings of classification systems are discussed in the following paragraphs. In Table 2, these bibliographic systems are evaluated for these metrics. The indicator ✓ shows the presence of metric value, × indicator represents that the system has no or minimal support for the mentioned metric, whereas and N/A is used for not applicable. In addition, each classification system has been evaluated and rated based on these metrics (Table 3), where Figure 1 graphically demonstrates the rankings and ratings of these classification systems.

Schemes / *Metrics*	**CC**	**UDC**	**DDC**	**LCC**	**ACM**	**IEEE**	**DMOZ**
Structural Complexity	✓	✓	×	×	×	×	×
Notational Brevity	×	×	✓	✓	✓	✓	N/A
Predefined Structure	✓	×	✓	✓	✓	✓	✓
Rules Complexity	✓	✓	×	✓	×	×	×
Theoretical Laws	✓	✓	✓	✓	×	×	×
Mnemonics	✓	✓	✓	✓	✓	✓	×
Hospitality	✓	✓	✓	✓	✓	✓	✓
Search Complexity	✓	✓	×	×	×	×	✓
Usability	✓	✓	✓	✓	✓	✓	×
Accuracy and Precision	✓	✓	✓	✓	✓	✓	×
Multilinguality	✓	✓	✓	✓	×	×	✓
Interoperability	×	✓	✓	✓	✓	✓	×
Semantic search	✓	✓	✓	✓	✓	✓	×
Bias in representation	✓	✓	✓	✓	✓	✓	×
Enumerative Structure	×	✓	✓	✓	×	×	×

Faceted Structure	✗	✓	✗	✗	✓	✓	✗
Faceted Search	✗	✓	✓	✓	✓	✓	✗
Consistency	✓	✓	✓	✓	✓	✓	✓
LOD Datasets	✗	✓	✓	✓	✓	✓	✓
LOV Support	✗	✗	✗	✗	✓	✗	✗
Platform	N/A	UDC consortium	OCLC	Library of Congress	ACM digital library	IEEE Xplore digital library	Open Directory Project
Warrants of classification	Literary Warrant (Giess, Wild, & McMahon, 2007)	Literary Warrant (Perles, 1995)	Literary and Scientific Warrant (Giess et al.,2007)	Literary and Scientific Warrant (Giess et al.,2007)	Scientific Research warrant	Scientific Research warrant	N/A

Table 2. Evaluation of Classification Schemes

The structural complexity means difficulties in using the structure and notations in classifying and describing a specific subject area. The metric will help us in selecting a classification scheme or system that is easy to use in classifying document collection by requiring short notations and simple rules. The notations and rules are complex in CC and UDC (Ranganathan, 1968). This complexity is because of the faceted structure in these classification schemes (Sukhmaneva, 1970). The structural complexity of CC is greater than that of UDC. UDC comes at second position in complexity as compared to CC. Because of its enumerative structure, LCC stands at third position, as it is lesser complex than CC and UDC. DDC is the simplest in this list because it is based on enumerative classification structure and on the principle of dividing universe of knowledge into defined classes. IEEE is more complex than ACM, whereas DMOZ is the least complex system. The classification system with greater structural complexity is ranked lower in the list. Therefore, based on this metric, the classification systems can be ranked as DMOZ, ACM, IEEE, DDC, LCC, UDC, and CC.

The notational brevity means how brief are the notations in describing and understanding the holdings with minimum number of symbols and minimal cognitive load. DDC uses well-organized short notations and their mnemonic value is also greater (Comaromi & Satija, 1983) (Hyman, 1980). LCC has notational brevity (Chan et al., 2016). UDC uses lengthy notations (Kaosar, 2008) as compared to DDC, whereas CC also uses lengthy and complex notations (Chatterjee, 2016). ACM notations are shorter than IEEE, whereas DMOZ do not use any notations at all. Using this metric, these classification systems can be ranked as ACM, IEEE, DDC, LCC, UDC, CC, and DMOZ at the last with no usage of symbols at all.

The predefined structure means that the classification scheme follows rigid pre-assumed subject categorization along with classification class marks. In this regard, UDC and LCC are enumerative

and impose subjectivity viewpoint of classification by following a predefined structure (Goh, Giess, McMahon, & Liu, 2009). Being faceted, CC arranges basic concepts in few predefined categories (Satija & Martínez-Ávila, 2015). DDC also has the predefined hierarchical structure of classification (Press, 2002) (Jonassen, 2004). Among these schemes, CC has minimal predefined structure because of using facets; UDC is both enumerative and analytico-synthetic. LCC is enumerative but possesses weaker predefined rules for the structural design. Because of the rigid enumerative hierarchies and predefined class structure, DDC comes at first position. DMOZ has the most rigid predefined structure as compared to that of IEEE and ACM. The classification system with most rigid and predefined structure is ranked lower, and therefore, the ranking could be CC, ACM, IEEE, UDC, DDC, LCC and DMOZ.

The complexity in rules determines the difficulty level in applying classification rules on knowledge artifacts. CC presents a complex set of rules and classification theory, which is comparatively difficult to implement and understand (Tennis, 2011). LCC is also complex ("Library of Congress Subject Headings: Pre- vs. Post-Coordination and Related Issues," March 15, 2007) in implementing Library of Congress Subject Headings (LCSH) in pre-coordinated subject strings. DDC's rules and principles are comprehensive and complete (Press, 2002) and easier than those of CC and LCC. UDC is also easy to understand and implement (Piros, 2014). ACM, IEEE, and DMOZ are simple to use and understand, and therefore, bears no such complexity. A classification system with greater complexity is ranked lower, therefore, based on this metric, the rankings could be ACM, IEEE, DMOZ are on top with similar rankings followed by UDC, DDC, LCC, and CC.

Theoretical laws are considered as a metric to analyze the foundations of classification systems to understand whether they are based on certain theoretical laws and principles of classification or not. UDC combines the enumerative and faceted approaches gathered from DDC and CC (Kaosar, 2008). The synthetic principle of UDC contributes to its widespread use but it is not enough at the intellectual level for making the relations between the subject facets (Kyle & Vickery, 1961). UDC lacks standard rules for its application for making facets, but there are rules for its structural representation (McIlwaine, 1997). Therefore, the structural and synthetic rules are good enough for its applicability but it should be refined further at the intellectual level. The theoretical laws of CC are based on the faceted approach of managing knowledge artifacts. CC has sound rules and principles, which include different postulates, laws, principles and canons (Batley, 2014) (Arashanapalai Neelameghan & Parthasarathy, 1997). On the other hand, LCC has weaker theoretical foundations. There also exists some intellectual and structural limitations due to its enumerative structure (San Segundo Manuel, 2008). DDC has the hierarchical and the enumerative structure which is based on the knowledge philosophy of hierarchical division (Hjorland, 1999). Because of the strong theoretical foundations, CC is at the top of this list, DDC is second because of its universal theory of knowledge division, UDC is third for being exploiting the theories of DDC and CC, LCC is at fourth position for comparatively weak theory of classification, whereas ACM, IEEE, and DMOZ present no or very limited theoretical laws or philosophical rules of classification.

The support for using mnemonics enables human classifiers to easily memorize the symbols and notations of classification scheme. The systematic and literal mnemonics are used in UDC (Satija, 2013) (Kaosar, 2008). The mnemonics are increased through mnemonic devices, which are described through the canons of mnemonics (Kaula, 1965). LCC uses literal mnemonics (Satija, 2013), whereas DDC uses systematic and literal mnemonics but its systematic mnemonics are not consistent (Satija, 2013). There are several seminal mnemonics in CC (Rahman & Ranganathan, 1962). These mnemonic devices increase mnemonics in CC, but the formation and length of the notations affects this mnemonic quality. ACM has greater support for mnemonics in comparison with IEEE, whereas DMOZ is the collection of web pages under specific categories. Based on this metric, the rankings of classification systems could be DDC, UDC, LCC, ACM, IEEE, CC, whereas DMOZ lacks in using any mnemonic devices or notations.

Hospitality means the ability of a classification scheme to incorporate new knowledge areas expressed in different multilingual contexts. Hospitality is present in UDC (Kaosar, 2008). CC is also hospitable for new subjects (De Grolier, 1962). LCC is hospitable for expressing the new subjects and knowledge areas (Satija, 2013). DDC is hospitable for new subject areas (Satija, 2013). By applying this metric, a classification scheme with faceted approach is naturally more hospitable than others. Therefore, CC is more hospitable and at the top in this list followed by UDC. DDC is at third position for being following enumerative approach. LCC is at fourth position because of it's of pure enumerative structure. IEEE and ACM are at fifth position by covering short span of knowledge areas, faceted structure, and efficient search. DMOZ is covering only web pages in already specified categories therefore it is at seventh position.

Search complexity measures the difficulty in searching artifacts using a classification scheme. It describes that which classification scheme is worth in searching a specific document. Search complexity is minimal in UDC because of its syntactic-analytico and enumerative nature (Kaosar, 2008), which can contribute in search applications in both Web based and in-house searching applications e.g., Online Public Access Catalog (OPAC). The theory and philosophy of CC is the trend setter for the knowledge management, resource discovery & access, however, according to (Raghavan, 2016) searching through CC is comparatively weaker than other bibliographic classification schemes. According to (Chan, 2000), LCC and LCSH have the potential to provide the ease in searching because of richer vocabulary for greater subject coverage, synonym and homograph capabilities, pre-coordinated system, browsing capability in multi-faceted structure, multilingual support and MARC format support with sematic interoperability. However, it is limited in providing ease in search & retrieval process, which include syntax and application rules complexity, lack of training for the personnel, and too lengthy and complex searching strings. DDC and LCC are aggregated in Classify[12] project initiated by OCLC. With the use of the Classify application, the search experience of the catalogers and patrons becomes much easier. Using this metric, DDC stands at the top with least complexity than LCC, UDC, and CC. IEEE is more complex than ACM and DMOZ. The classification scheme with less search complexity will be ranked higher.

[12] http://www.oclc.org/research/themes/data-science/classify.html

Therefore, ACM and IEEE, DDC, and LCC stand first with least search complexity followed by UDC, and CC. DMOZ stands at the last position with greater search complexity having loose boundaries of categorization.

Usability analyzes the difficulty in using a classification scheme for classifying and searching documents. This metric defines the ease of learning and effective usage. Usability measures user satisfaction, user understanding of the system, and precision with minimal recall in lesser amount of time (Singapore, 2016). OCLC has included structural changes to improve usability and simplify classification tasks ("Dewey Services: Dewey Decimal Classification System,"). The Classify[13] project aims at finding books through a web interface, which is easy to use and understand by using DDC and LCC. UDC is extensively used in Web-based search and retrieval applications (Kaosar, 2008). This classification scheme is used in several institutions' OPAC systems ("Library OPACs containing UDC codes,"). The UDC notations are supportive for the usability (Slavic-Overfield, 2005). However, the user interface of these OPAC search systems could be further improved (Slavic, 2006) (Pollitt, 1998) (Schallier, 2005). CC is the source of inspiration and a standardized model for the usability of faceted structure of bibliographic classification in the electronic and web based environments (Thelwall, 2009). In (Rosenfeld & Morville, 2002), the philosophy and methodology of CC is considered at the abstract and theoretical level. This assessment of CC leads us to the argument that the faceted structure is supportive in precise retrieval with a considerably high cognitive work at the user end as compared to DDC and LCC because of their simple enumerative structures. Library of Congress uses LCC in its catalog[14] and Classification Web[15] applications. These applications are exploiting LCSHs and LCC in user friendly manner. By looking at the usability aspect of these classification schemes, the ranking through this metrics appears as DDC is at the top for its easy enumerative structure and notational simplicity along with easy to use Web applications. LCC is at second position because of its enumerative structure and adoptability in web applications. Being enumerative and faceted, UDC stands at the third position. CC for being a pure faceted scheme with complex notations and rules, is ranked at the fourth position. Similarly, IEEE and ACM are faceted and easy to use, and therefore, share the first position with DDC. DMOZ with loose boundaries of categorization is least usable with limited browsing and search.

The accuracy and precision metric measures how accurate and precise a classification system can identify the exact locations of the holdings in the given knowledge space. UDC shows accuracy and precision in finding the required knowledge artifact (Kaosar, 2008). The accuracy and precision of CC gets compromised as its lengthy notations introduces complexity in searching and discovering documents (Satija, 2015). LCC and DDC were researched for accuracy and precision by using a prototype model (Gnoli, Pusterla, Bendiscioli, & Recinella, 2016) for automatic text classification of electronic documents using classification metadata of library holdings from LCC and DDC

[13] http://classify.oclc.org/classify2/
[14] https://catalog.loc.gov/vwebv/searchBasic
[15] https://www.loc.gov/cds/classweb/classwebfeatures.html

datasets. It was observed that for precision, there is a need for increasing DDC and LCC bibliographic data on the Web, introducing searching capabilities for bibliographic data at the micro level of any document, and increasing the efficiency of user interfaces for navigation using DDC-based browsing structure (Joorabchi & Mahdi, 2009) (Joorabchi & Mahdi, 2011). Therefore, CC because of the pure faceted approach has high-level precision in search and resource discovery. UDC stands second because of being enumerative and enumerative and analytico-synthetic. DDC is at the third position as OCLC maintains and updates its structure regularly along with state-of-the-art search applications. LCC shares the third position with DDC, being regularly updated and maintained by Library of Congress for precision in their search application. IEEE and ACM also show greater precision in their search & retrieval, and therefore, share the third position with DDC and LCC. DMOZ are the manually created and updated categories of web pages, having limited keyword search with very low precision.

In connection to the evaluation framework, multilinguality means to classify and describe the knowledge artifacts written and expressed in variety of natural languages and the availability of any classification scheme in different natural languages. DMOZ supports 72 different languages of the world and therefore stays at the top. UDC is multilingual by supporting French, Portuguese, Spanish and Russian (Slavic, 2008) (Koch & Day, 1997) and has been translated into languages ("Universal Decimal Classification summary," 2017). LCC supports works in 19 language subclasses ("Library of Congress Classification Outline: Class P - Language and Literature,") including German, Slavic, Oriental Languages and Roman languages etc. The translations of DDC support to localize this scheme for different languages of the world (Vizine-Goetz, 2009). DDC is translated in 30 different languages but covers different languages in only seven classes i.e., from 420 to 490 class number ("Dewey Decimal Classification summaries,"). CC shows minimal multilingual support because of its sub-continental origin (A Neelameghan & Lalitha, 2013; Raghavan, 2016). ACM and IEEE are in English languages only and therefore, show no multilinguality at all. Using this metric, we can conclude that DMOZ is at the first position, followed by UDC, DDC, LCC, and then CC.

Consistency measures the level of uniformity in classification system to classify subjects. According to (Batty, 1967), in the earlier stages, CC shows no consistency but by the addition of consistency cannons, it has gradually become consistent. LCC seems less consistent in expressing different subjects areas (Madge, 2011). DDC and LCC were found short of defining and classifying religious holdings especially Jewish contents. These schemes also show biasness towards different religious and regional contents (Maddaford & Briefing). Although DDC is a little bit inconsistent, still it can classify complex subjects (Gnoli et al., 2016). UDC also shows inconsistency, which can be sorted out by introducing specific UDC classes to database in online system (Kaosar, 2008). DDC shows comparatively greater consistency in classifying new subjects with constant uniformity; CC is ranked second because of the introduction of canons of consistency. LCC and UDC are ranked at third position. For being only limited to the scientific research articles, IEEE

and ACM are at fourth position. DMOZ stands at fifth position due to its loose boundaries of categorization.

The interoperability determines how much a given classification scheme is interoperable in expressing its classification artifacts with other schemes. UDC is interoperable (Koch & Day, 1997) and supports integration with other systems. CC, because of its sub-continental origin, shows limited interoperability (A Neelameghan & Lalitha, 2013) (Raghavan, 2016). LCC shows interoperability by being capable to map with DDC (Vizine-Goetz, 2009). The interoperability and multilinguality of DDC enables it to map with other classification schemes (Vizine-Goetz, 2009). IEEE, ACM and DMOZ datasets are interoperable with other web applications. Based on this metric, DDC, LCC, UDC, ACM, IEEE and DMOZ are standing at first position because of the presence of their interoperability and data harvesting protocols and ontologies in the digital environment. DMOZ stands at the second position because of limited interoperability. CC provides only philosophical and theoretical model but we found no practical web-based application so it is not included in this list.

By enabling semantic search, a classification scheme can proactively respond to information seekers using its faceted structure. UDC, because of its semantic structure (Slavic, 2008), contains semantic search capability. The classification theory and philosophy of CC provides the basis for classification ontology development (Panigrahi & Prasad, 2005), which makes obvious its capability of semantic search and inference. LCC supports semantic search through LOD support, semantically enabled LCSH and authority control files ("LC Linked Data Service: Authorities and Vocabularies,") (Harper & Tillett, 2007). DDC also contains semantic features (Green, 2015), which can be utilized in the semantic search applications. Therefore, it can be concluded that semantic search is also supported by DDC. This metric can be better analyzed in the digital environment and especially through analyzing these bibliographic classifications for their ontologies. LCC could be ranked first because of its expressive ontology with efficient semantic search application. DDC is at second position, because of efficient search but limited usage of its ontology. ACM is at third position because of its expressive ontology and efficient search but limited coverage to scientific domain. IEEE is at fourth position because of its faceted semantic search. UDC comes at fifth position because of its ontological presence but with limited usage. CC has no application in the digital environment, which could demonstrate its capability for semantic search, although it provides the basis for the semantic level for all bibliographic classification systems. DMOZ lacks in semantic search, where it is only based on keywords.

Bias in subject representation means inclination for or against some subjects which results in unfair, partial negligence or fully ignoring any subject. DDC and LCC are biased in representing different knowledge and regional information, e.g., Anglo-American bias (Tomren, 2003), while UDC is biased towards European culture (Fandino, 2008). CC is biased towards different knowledge areas (Satija & Singh, 2010). A classification system with least biasness is ranked higher. Therefore, in this connection, DMOZ is ranked higher for showing no/least biasness; CC is ranked second because of the presence of acute biasness followed by DDC showing comparatively

less biasness towards religion and regional subjects. LCC comes at the fourth position followed by IEEE and ACM that show greater biasness towards certain domains.

Enumerative structure exhibits the rigid hierarchies. LCC is enumerative (Goh et al., 2009; Perles, 1995) (Bryant, October 4, 1993). UDC is nearly enumerative and faceted (Kaosar, 2008) (Bryant, October 4, 1993) and DDC is both analytico-synthetic and enumerative (Hallows, 2014). CC is faceted (Chatterjee, 2016; Dawson, Brown, & Broughton, 2006). By comparing these systems, LCC fully supports enumerative structure, and then comes DDC, whereas UDC is nearly enumerative and CC shows no enumerative structure at all. LCC and DDC are enumerative. The trend is towards semantic and faceted structure, and therefore, enumerative structure in classification systems is not a desirable characteristic. Therefore, the system with enumerative nature will be ranked lower. Based on this metric, CC and DMOZ are least enumerative and therefore, ranked higher, followed by IEEE and ACM at the second position, then UDC at the third position, while DDC and LCC at the last.

The faceted structure means the semantically interlinked structure of categories, which can be merged and combined to generate an expression for existing or new concepts (Svenonius, 2000). CC is faceted (Chatterjee, 2016; Dawson et al., 2006). UDC is analytico-synthetic (Kaosar, 2008) and follows the faceted method of CC using different connecting symbols in mixed notations and using subject facets including time and space (Chatterjee, 2016). IEEE and ACM possess faceted structures. DMOZ has only hierarchical structure and predefined categories. Based on this metric we rank CC first, UDC second, ACM and IEEE third while DDC and LCC are enumerative structures, and therefore, cannot be included in the list.

Faceted search means to navigate or browse through the faceted structure of a faceted classification scheme. Faceted search is also applied by selecting different ranges and choices from different facets that are given by any faceted system to search the required contents. It is different from search complexity in the sense that it looks at the pattern and criteria of search that exist in any classification scheme either in there OPACs or web applications. The theory and philosophy of CC supports faceted search & browsing economically (Kong, 2016), however, to the best of our knowledge, no real-world application demonstrates its usefulness. UDC is based on the faceted approach, which supports faceted search (Tunkelang, 2009). LCC supports faceted search with the help of LCSH (McGrath, 2007). LCC also provides faceted search through the Faceted Application of Subject Terminology (FAST) application ("Faceted Application of Subject Terminology," 2017). DDC provides the faceted search through the OCLC Classify[16] application. Using this metric, these classification schemes can be ranked as DDC at first position because DDC is adopting the faceted approach along with its native enumerative nature and state-of-the-art web based search applications developed by OCLC. LCC is at second position because of its web based search applications and its adaptation of comparatively restricted faceted approach. IEEE, for providing extensive choice of searching patterns, stands at the third position. ACM has poly-hierarchical and

[16] http://classify.oclc.org/classify2/

multi-faceted classification structure along with robust search mechanism; therefore, it is on fourth position in this list. There are very limited faceted search applications of UDC and therefore it stands at fifth position. DMOZ has hierarchical structure in which the required element can be accessed through a keyword search. Therefore, it is not providing any faceted search. CC has no search applications that could confirm its support for the faceted search.

LOD datasets means the availability of datasets of a given classification system on LOD cloud. Among our choice of well-known classification systems UDC, LCC, DDC, IEEE, ACM and DMOZ have datasets in the LOD cloud whereas CC has no such datasets. The definitions of classes and properties are gathered in Linked Data Vocabularies (LOV), which are used for describing different types of objects used in LOD cloud. These definitions of different things provide vocabularies for linking the linked data (Foundation, 2017). CC, UDC, DDC, LCC, IEEE and DMOZ have no LOV, whereas ACM has LOV vocabularies.

The metric "platforms" in the evaluation framework, considers the applicability of a given classification system in real-world web applications and other digital environments. In this regard, UDC is supported by UDC consortium, DDC by OCLC, LCC by Library of Congress, ACM by ACM digital library, IEEE by IEEE Xplore digital library, and DMOZ by Open Directory Project. To the best of our knowledge, CC has not been used by any of the online applications.

Ranking	Structural Complexity	Notational brevity	Predefined Structure	Rules Complexity	Theoretical Laws	Mnemonics	Hospitality	Search Complexity	Usability	Precision and [illegible]	Multilinguility	Interoperability	Semantic Search	Biasness	Enumerative Structure	Faceted Structure	Faceted Search	Consistency	LOD datasets	LOV Support	Average Ranking
CC	1	2	7	1	5	2	6	2	2	4	2	1	7	5	4	4	1	4	1	1	3.1
UDC	2	3	4	4	3	6	5	3	3	3	5	3	3	3	2	3	2	5	2	1	3.25
DDC	4	5	3	3	4	7	4	4	5	2	4	3	6	4	1	1	6	3	2	1	3.6
LCC	3	4	2	2	2	5	3	4	4	2	3	3	2	1	1	1	5	3	2	1	2.65
ACM	6	7	6	5	1	4	2	4	5	2	1	3	5	2	3	2	3	2	2	1	3.3
IEEE	5	6	5	5	1	3	2	4	5	2	1	3	4	2	3	2	4	2	2	2	3.15
DMOZ	7	1	1	5	1	1	1	1	1	1	6	2	1	6	4	1	1	1	2	1	2.25

Table 3. Ranking and Average Ranking of Classification Schemes

The warrants of classification work as authoritative acts for classificationists to perform the cognitive practice for designing the classes and concepts in the classification system, their structural properties and then putting subjects in the specified classes (Beghtol, 1986). CC and

UDC use literary warrant; DDC and LCC use literary & scientific warrants. ACM and IEEE use scientific research warrant, while DMOZ exhibits no warrant of classification.

In the above paragraphs, we compared and evaluated the selected classification system using the evaluation metrics (shown in Table 2), and discussed how these systems can be ranked based on a given evaluation metric. However, to give a holistic view of this comparison and evaluation, we introduce a ranking score or levels ranging from 1 (meaning low ranking, not applicable, or not available) to 7 (meaning high ranking) in how a classification scheme is best among its counterparts in the list. It is also the case that for a given metric, multiple systems may belong to the same ranking level. By assigning these ranking levels, Table 3 compares these systems based on 20 metrics by excluding platforms and warrants of classification. Table 3 also reports the average ranking of these classification systems, showing DDC at top with average ranking of 3.6, followed by ACM = 3.3, and UDC = 3.25. It can be concluded that DDC and UDC are among the best classification schemes for describing printed as well as digital collections, whereas ACM is best for classifying digital collections belonging to Computer Science domain. However, ACM classification system can be extended to include other domains as well. Figure 1 illustrates graphically the comparison and evaluation of these systems.

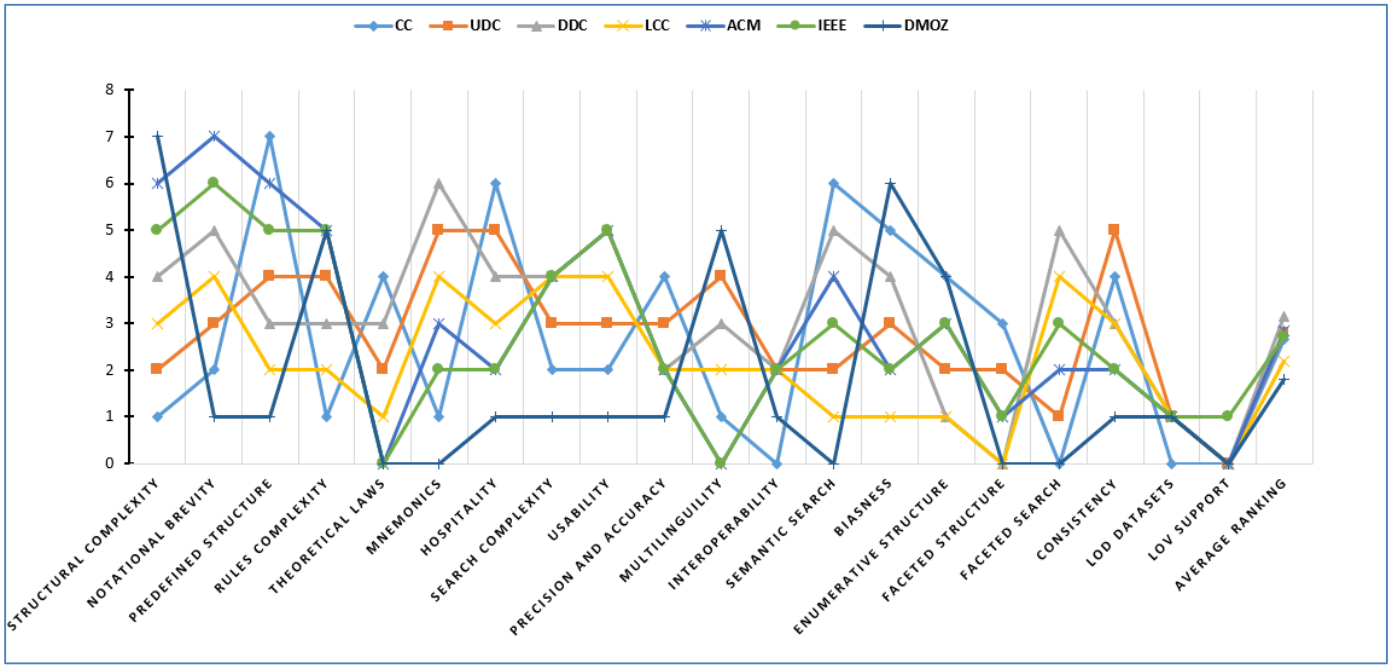

Figure 1. Comparison and Ranking of Classification Systems

Table 4 presents the state-of-the-art bibliographic classification ontologies including Bibliographic ontology, LCC ontology, DDC ontology, UDC ontology, and DMOZ ontology. Some of these ontologies were designed specifically for certain targeted applications e.g., ACM ontology for ACM digital library, and LCC ontology for Library of Congress etc., whereas others have multiple usage scenarios and have been used by several applications. An example of such general-purpose

doi:10.6017/ital.v36i3.8930

bibliographic classification ontology is the Bibliographic Ontology[17], which is used by several bibliographic services and digital libraries e.g., digital object identifier (DOI), Zotero, and Library of Congress Classification Number (LCCN) permalink service (Giasson, 2012). This evaluation framework compares these ontologies based on their size (in terms of number of classes), usage in the state-of-the-art applications, LOD support, the availability of datasets on datahub[18], and LOV support. By looking at Table 3, ACM show more comprehensiveness in terms of number of classes, triples and LOV support.

Classification and Categorization Ontologies	No. of classes	Applications	LOD datasets	LOD datasets triples	LOV support
Bibliographic ontology[19]	69	Library of Congress and BibBase	✓	200000	✓
LCC ontology[20]	40+	Library of Congress	✓	Not Given	×
DDC ontology[21]	20+	OCLC	✓	402288	×
UDC ontology[22]	2,600	UDC[23]	✓	69,000	×
ACM ontology[24]	1469	ACM	✓	12402336	✓
IEEE LOM metadata ontology **(Casali, Deco, Romano, & Tomé, 2013)**	9	IEEE[25] Xplore digital library	✓	91564	✓
DMOZ ontology[26]	Not given	Open Directory Project	✓	Not given	×

Table 4. Comparison of classification and categorization ontologies

[18] https://datahub.io
[19] http://purl.org/ontology/bibo
[20] http://id.loc.gov/
[21] http://dewey.info/
[22] http://udcdata.info/
[23] http://udcdata.info/
[24] http://dl.acm.org/ccs/ccs.cfm
[25] http://ieee.rkbexplorer.com/id/
[26] https://www.dmoz.org/rdf.html

Issues & Challenges in Classification Research

Although bibliographic classification has been practiced since the use of books and the inception of library & information science practices, further research & development efforts are required for to meet the classification needs of the digital age. Especially, with the arrival of digital holdings, researchers face several issues and challenges. For example, automatic text classification performs categorization of resources using ordinary metrics including TF-IDF and classification in its true sense is yet to be achieved (Yi, 2006). To handle the issue, text classification is also carried out through semantic indexing but its accuracy and precision are yet to be achieved. Semantic and structural relationships among different parts of text corpus is still at infant level and has not been exploited to their fullest so that these can be used in text classification in more meaningful ways. Other challenges in text classification include handling huge data resulted by applying a classification scheme, dynamism in classification, and structure dissimilarity among classification schemes although they agree upon subject as the primary characteristic.

The biasness in DDC and LCC needs to be resolved. Several revisions and proposals are put forward for addressing the problem of systematic knowledge organization and searching through natural language terms (Miksa, 2007). There are various issues regarding the structural updates, search & retrieval criteria, and visualization (Slavic-Overfield, 2005).

There are two main challenges for the application of the bibliographic classification principles in classifying the Web. First, the principles of the bibliographic classification are formulated for the printed documents, which should also be applicable to digital collections. For addressing these challenges, there is need to apply and modify bibliographic classification principles in digital environments. Second, it is required to exploit hidden hierarchies and concepts to be better classified by the principles of bibliographic classification for precise discovery, search and retrieval (J. Mai, 2004).

The issue of dependent process of classification of any object per predefined criteria and principles is important to address for finding a place in this age of search engines. This issue can be tailored by the principles of classification, so that the conventional principles are modified to consider the purpose of classification and domain of objects. For this issue semantic web and ontologies can play a vital role in bibliographic classification, which can provide independent classification of the bibliographic classification predefined theories (Hjørland, 2012).

The issue of heterogeneity conflicts, which arise because of the inconsistencies and structural divergences, are the challenges for the semantic interoperability. Semantic interoperability can be brought into the bibliographic records inside the bibliographic system and across the systems through different phases of interlinking, evaluation, analysis, remodeling & conversion for analyzing, and restructuring the bibliographic data (Tallerås, 2013).

Bibliographic data is in multi-format, multi-topical, multi-lingual and multi-targeted. For tackling these issues, the bibliographic data must be made mutually interoperable for making it

interlinked, searchable, and presented in a harmonized way across the boundaries of the datasets and data silos. The interoperability problem arises at the syntactic level for making consistent the character sets, notations, data formats and records in different systems. The interoperability problem is also arising at the semantic level because of the difference in data interpretations and difference in vocabularies, and precision levels in data encoding. Publishing, collecting and maintenance of bibliographic data by multi organizations through own established standards and best practices in Web 2.0 (Hyvönen, 2012). With these problems in hand, the transition of this data from syntactical Web to Semantic Web is a challenge for bringing the uniformity in records that are generated by diverse sources, encoded in multi-bibliographic systems, cross bibliographic systems interoperability, the visualization of bibliographic data accordingly as per need for different contexts. For addressing these problems there is a need for coordination and collaboration between bibliographic data publishers and the technical developers of the web applications (Hyvönen, 2012).

There is variety of metadata standards and schemas for defining, managing, resource discovery, search & retrieval, preserving, mapping, cross-walking, integrity, accuracy, and authenticity of metadata and bibliographic data. But for these tasks to be handled with great simplicity, semantic richness and accuracy, a universal all in one metadata format and schema is the need of the day (Ramesh, Vivekavardhan, & Bharathi, 2015) to get out of this jungle of standards (Gartner, 2016). This way, the metadata publishers and managers could get relieved and the job will become economic in terms of time, management, and search & retrieval.

Three main tasks were set in Semantic Publishing challenges 2015. These tasks are: (i) extracting data on workshops' quality indicators; (ii) extracting data on affiliations, citations, funding; and (iii) interlinking. Several challenges were faced while fulfilling these tasks. These tasks are being fulfilled through a proposed solution, which is composed of a text mining pipeline, LODeXporter and Named Entity Recognition (NER) for named entities extraction form text and linking them to resources on the LOD cloud (Sateli & Witte, 2015).

In (Peroni, 2012) three main issues of semantic publishing are addressed which are: lack of document publishing universal metadata schemas according to publishing vocabulary, lacking of efficient user interface that are based on models and theories of semantic publishing, and there is a need for a tool that semantically link and describe document text. These issues require the urgent need for comprehensive ontologies for document publishing domain.

(Ferrara & Salini, 2012) tossed 10 challenges for multiple dimensions of data in terms of bibliographic analysis. These challenges are: (i) analyzing bibliographic data in a multidimensional pattern; (ii) discovering and integrating data coming from diverse sources; (iii) detecting multiple references to the same item and cleaning, normalizing, and disambiguating bibliographic data records; (iv) analyzing multidimensional nature of bibliographic data through multivariate analysis for aggregating the data; (v) comparing different elements of bibliographic data and its ranking accordingly, (vi) aggregating indexes of different nature with respect to

different parameters, dimensions, and elements of bibliographic data; (vii) dealing with multiple indexes for the same item with different values coming from different sources; (viii) extracting and indexing textual information from text corpus in support of text mining; (ix) analyzing textual data topic-wise and describing these topics for research and learning process and tracing different trends; and (x) combining multidimensional information for finding trends in bibliographic data collection.

Bibliographic classification systems are being incorporated in LOD. In Dewey.info[27], a prototype version of DDC is designed for linking its dataset in linked data cloud. The intention is to provide a platform for DDC data on the Web having summaries of top 3 levels of classification order of DDC 22nd edition in 11 different language encoded in RDF/SKOS, having actionable URIs for every class, representation for machines is in RDF, and for humans in XHTML+RDFs, and serialization available in formats of RDF/XML, Turtle and JSON, and with SPARQL endpoint. (OCLC 2011; Mitchell and Panzer 2013). However, this version of DDC on LOD cloud is still at infant stage to cover different subjects and to be widely used in generating and creating documents metadata.

Library of Congress Linked Data service provides access to commonly used standards and vocabularies developed by Library of Congress. This includes data values, controlled vocabularies, and preservation vocabularies which are part of this service. This service provides access to LCSH, LCC name authority files, LCC[28], LC children's subject headings, LC genre/form terms, thesaurus for graphic materials, MARC relators, MARC countries, MARC geographic areas, MARC languages, ISO639-1 languages, ISO639-2 languages, ISO639-5 languages, extended date/time format, preservation events, and preservation level role and cryptographic hash functions. The authorities and vocabularies currently included in this service are listed on the Linked Data service (Library of Congress 2014). However, it lacks in vocabularies for supporting PREMIS, MARC, MODS, METS, and MIX.

As presented in Section 2, several ontologies have been developed for describing and sharing knowledge about bibliographic classification. However, the available ontologies are limited in several ways e.g., these ontologies are not the complete clones of classification schemes of which they are deemed to be ontologies and they also not mature enough in terms of metadata collection. In addition, these ontologies still couldn't break the cross-classification scheme metadata collection barriers i.e., they are not interoperable enough to harvest the metadata across bibliographic ontology system. Therefore, further initiatives are required to develop matured bibliographic ontologies which fully clone bibliographic schemes that are in practical use and have strong theoretical ground. These ontologies must be interoperable and sharing metadata collection with other bibliographic ontologies. In this way in future we can have ontology-based general bibliographic classification system by fusion of the new and existing bibliographic ontologies for better management of the knowledge artifacts.

[27] https://datahub.io/dataset/dewey_decimal_classification

doi:10.6017/ital.v36i3.8930

CONCLUSIONS

With the arrival of digital collections, new challenges of preservation, curation as well as resource discovery & access (retrieval) have emerged that needs proper attention, where classification schemes and ontologies can play a significant role. By comparing and evaluating the available bibliographic classification and categorization systems it is concluded that currently DDC is the best classification system followed by UDC, and ACM. The bibliographic classification ontologies are limited in one way or the other e.g., some of these are comprehensive like UDC and ACM but lack support for LOD and LOV etc., while others support these later aspects but lack comprehensiveness. Keeping in view the available bibliographic classification ontologies and their limitations, we recommend that a universal bibliographic classification ontology should be developed by using the classes from the available ontologies and providing support in terms of availability of datasets, support for interoperability, LOD, and Linked data vocabularies.

For developing a more meaningful classification system, equally applicable to digital environments, it is necessary to consider the book structural semantics such as table of contents, headings, chapters, sections, subsections, figures, algorithms, mathematical equations, quotations etc., and the logical connections in contents (Khusro & Ullah, 2016; I. Ullah & Khusro, 2016) as well as about the book information i.e., the bibliographic details of the holdings. To meet, the former requirement, a comprehensive ontology like BookOnt (A. Ullah, Ullah, Khusro, & Ali, 2016) could be used, which can be mapped with any bibliographic ontology like e.g., Bibliographic Ontology[29]. However, as the evaluation frameworks suggest, DDC, UDC, and ACM Classification System should be exploited in designing such a general-purpose classification system.

REFERENCES

The 2012 ACM Computing Classification System. Retrieved March 20, 2017, from http://www.acm.org/about/class/2012

About Universal Decimal Classification (UDC). Retrieved March 21, 2017, from http://www.udcc.org/index.php/site/page?view=about

Albrechtsen, H. (2000). Who wants yesterday's classifications? Information science perspectives on classification schemes in common information spaces. In K. Schmidt (Ed.), *Papers*. Technical University of Denmark, Center for Tele-Information.

Batley, S. (2014). *Classification in Theory and Practice*. Oxford: Chandos Publishing.

Batty, C. D. (1967). *An introduction to colon classification*: Archon Books.

Beghtol, C. (1986). Semantic validity: concepts of warrant in bibliographic classification systems. *Library resources & technical services, 30*(2), 109-125.

[29] http://bibliontology.com/#

Beghtol, C. (1998). Knowledge domains: multidisciplinarity and bibliographic classification systems. *Knowledge Organization, 25*(1-2), 1-12.

Beghtol, C. (2001). Relationships in classificatory structure and meaning *Relationships in the organization of knowledge* (pp. 99-113): Springer.

Berners-Lee, T. (2006, June 18, 2009). Linked data. *Design Issues.* Retrieved March 21, 2017, from https://www.w3.org/DesignIssues/LinkedData.html

Betts, T., Milosavljevic, M., & Oberlander, J. (2007). The utility of information extraction in the classification of books *Advances in Information Retrieval* (pp. 295-306): Springer.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205-227.

Boykin, J. (2016). Assessing DMOZ: A Quality Review. Retrieved 14-03-2016, 2016, from https://www.seochat.com/c/a/search-engine-news/assessing-dmoz-a-quality-review/

Bryant, B. (October 4, 1993). 'Numbers You Can Count On' Dewey Decimal Classification Is Maintained at LC. *Library of Congress Information Bulletin, 52*(18). http://www.loc.gov/loc/lcib/93/9318/count.html

Buchanan, B. (1979). Theory of library classification.

Campbell, D. G. (2002). *Centripetal and Centrifugal Forces in Bibliographic Classification Research.* Paper presented at the ASIS SIG/CR Classification Research Workshop.

Casali, A., Deco, C., Romano, A., & Tomé, G. (2013). An assistant for loading learning object metadata: An ontology based approach.

Chan, L. M. (2000). Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources: Issues and Challenges.

Chan, L. M., Intner, S. S., & Weihs, J. (2016). *Guide to the Library of Congress classification*: ABC-CLIO.

Chapman, J. W., Reynolds, D., & Shreeves, S. A. (2009). Repository metadata: approaches and challenges. *Cataloging & classification quarterly, 47*(3-4), 309-325.

Chatterjee, A. (2016). Universal Decimal Classification and Colon Classification: Their mutual impact. *Annals of Library and Information Studies (ALIS), 62*(4), 226-230.

Cliff, P. (2008). JISC-Repositories: Subject Classification Thread Summary.

Comaromi, J. P., & Satija, M. P. (1983). *Brevity of notation in Dewey decimal classification*: Metropolitan.

Dawson, A., Brown, D., & Broughton, V. (2006). *The need for a faceted classification as the basis of all methods of information retrieval.* Paper presented at the Aslib proceedings.

De Grolier, E. (1962). A study of general categories applicable to classification and coding in documentation.

Dewey Decimal Classification summaries. Retrieved March 21, 2017, from https://www.oclc.org/en/dewey/features/summaries.html

Dewey Services: Dewey Decimal Classification System. Retrieved March 20, 2017, from https://www.oclc.org/content/dam/oclc/services/brochures/211422usb_dewey_services.pdf

Dorji, T. C., Atlam, E.-s., Yata, S., Fuketa, M., Morita, K., & Aoe, J.-i. (2011). Extraction, selection and ranking of Field Association (FA) Terms from domain-specific corpora for building a comprehensive FA terms dictionary. *Knowledge and Information Systems, 27*(1), 141-161. doi: 10.1007/s10115-010-0296-x

Dousa, T. M. (2009). Evolutionary order in the classification theories of CA Cutter & EC Richardson: its nature and limits.

Encyclopedia, N. W. (1 August 2014). Library classification. 2017, from http://www.newworldencyclopedia.org/entry/Library_classification

Faceted Application of Subject Terminology. (2017). Retrieved March 21, 2017, from http://www.oclc.org/research/themes/data-science/fast.html

Fandino, M. (2008). UDC or DDC: a note about the suitable choice for the National Library of Liechtenstein. *Extensions and Corrections to the UDC*.

Ferrara, A., & Salini, S. (2012). Ten challenges in modeling bibliographic data for bibliometric analysis. *Scientometrics, 93*(3), 765-785.

Foundation, O. K. (2017). About LOV. from http://lov.okfn.org/dataset/lov/about

Francu, V. (2007). *Multilingual access to information using an intermediate language: Proefschrift voorgelegd tot het behalen van de graad van doctor in de Taal-en Letterkunde aan de Universiteit Antwerpen.*

Gartner, R. (2016). *Metadata*: Springer.

Giasson, B. D. A. F. (2012). Projects using BIBO. from http://www.bibliontology.com/projects.html

Giess, M. D., Wild, P., & McMahon, C. (2007). *The use of faceted classification in the organisation of engineering design documents.* Paper presented at the Proceedings of the International Conference on Engineering Design 2007.

Gilchrist, A. (2015). Reflections on Knowledge, Communication and Knowledge Organization. *Knowledge Organization, 42*(6), 456-469.

Giunchiglia, F., Marchese, M., & Zaihrayeu, I. (2007). Encoding classifications into lightweight ontologies *Journal on data semantics VIII* (pp. 57-81): Springer.

Gnoli, C., Merli, G., Pavan, G., Bernuzzi, E., & Priano, M. (2008). Freely faceted classification for a Web-based bibliographic archive: the BioAcoustic Reference Database.

Gnoli, C., Pusterla, L., Bendiscioli, A., & Recinella, C. (2016). Classification for collections mapping and query expansion.

Goh, Y. M., Giess, M., McMahon, C., & Liu, Y. (2009). From faceted classification to knowledge discovery of semi-structured text records *Foundations of Computational, IntelligenceVolume 6* (pp. 151-169): Springer.

Green, R. (2015, October 29-30, 2015). *Relational aspects of subject authority control: the contributions of classificatory structure.* Paper presented at the Proceedings of the International UDC Seminar 2015 Classification & authority control Expanding resource discovery, Lisbon.

Hallows, K. M. (2014). It's All Enumerative: Reconsidering Library of Congress Classification in US Law Libraries. *Law Libr. J., 106*, 85.

Harper, C. A., & Tillett, B. B. (2007). Library of Congress controlled vocabularies and their application to the Semantic Web. *Cataloging & classification quarterly, 43*(3-4), 47-68.

Hjorland, B. (1999). The DDC, the universe of knowledge, and the post-modern library. *Journal of the Association for Information Science and Technology, 50*(5), 475.

Hjørland, B. (2007). Semantics and knowledge organization. *Annual review of information science and technology, 41*(1), 367-405.

Hjørland, B. (2008). Core classification theory: a reply to Szostak. *Journal of Documentation, 64*(3), 333-342.

Hjørland, B. (2012). Is classification necessary after Google? *Journal of Documentation, 68*(3), 299-317.

Hjørland, B. (2013). Theories of knowledge organization—theories of knowledge: Keynote March 19, 2013. 13th Meeting of the German ISKO in Potsdam. *Knowledge Organization, 40*(3), 169-181.

Hjørland, B. (2016). Subject (of documents). *Knowledge Organization, 44*(1), 55-64.

Hyman, R. J. (1980). Shelf classification research: past, present--future? *Occasional papers (University of Illinois at Urbana-Champaign. Graduate School of Library Science); no. 146 (Nov. 1980)*.

Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the semantic web. *Synthesis Lectures on the Semantic Web: Theory and Technology, 2*(1), 1-159.

Jacob, E. K. (2004). Classification and categorization: a difference that makes a difference. *Library trends, 52*(3), 515.

Jonassen, D. H. (2004). *Handbook of research on educational communications and technology*: Taylor & Francis.

Jones, K. S. (1970). Some thoughts on classification for retrieval. *Journal of Documentation, 26*(2), 89-101.

Joorabchi, A., & Mahdi, A. E. (2009). *Leveraging the legacy of conventional libraries for organizing digital libraries.* Paper presented at the International Conference on Theory and Practice of Digital Libraries.

Joorabchi, A., & Mahdi, A. E. (2011). An unsupervised approach to automatic classification of scientific literature utilizing bibliographic metadata. *J. Inf. Sci., 37*(5), 499-514. doi: 10.1177/0165551511417785

Kaosar, A. (2008). Merit & Demerit of using Universal Decimal Classification on the Internet.

Kaula, P. (1965). Colon Classification: Genesis and Development. *Library Science Today. Ranganathan's Festschrift, 1*, 87-93.

Khusro, S., & Ullah, I. (2016). *Towards a semantic book search engine.* Paper presented at the 2016 International Conference on Open Source Systems & Technologies (ICOSST'16), Lahore, Pakistan.

Koch, T., & Day, M. (1997). DESIRE - Development of a European Service for Information on Research and Education.

Koch, T., Day, M., Brümmer, A., Hiom, D., Peereboom, M., Poulter, A., & Worsfold, E. (1997). The role of classification schemes in Internet resource description and discovery. *Work Package, 3*.

Kong, W. (2016). *Extending Faceted Search to the Open-Domain Web.* University of Massachusetts Amherst.

Koshman, S. (1993). Categorization and classification revisited: a review of concept in library science and cognitive psychology. *Current Studies in Librarianship Spring/Fall, 26*.

Kwaśnik, B. H., & Rubin, V. L. (2003). Stretching conceptual structures in classifications across languages and cultures. *Cataloging & classification quarterly, 37*(1-2), 33-47.

Kyle, B., & Vickery, B. C. (1961). *The Universal Decimal Classification: present position and future developments*: Unesco.

LC Linked Data Service: Authorities and Vocabularies. Retrieved 28 Feb 2017, 2017, from http://id.loc.gov

Lee, H.-L. (2012). Epistemic foundation of bibliographic classification in early China: A Ru classicist perspective. *Journal of Documentation, 68*(3), 378-401.

Library of Congress Classification. (2014, 10/1/2014). Retrieved March 20, 2017, from https://www.loc.gov/catdir/cpso/lcc.html

Library of Congress Classification Outline: Class P - Language and Literature. [Press release]. Retrieved from https://www.loc.gov/aba/cataloging/classification/lcco/lcco_p.pdf

. Library of Congress Subject Headings: Pre- vs. Post-Coordination and Related Issues. (March 15, 2007) *Report for Beacher Wiggins, Director, Acquisitions & Bibliographic Access Directorate, Library Services, Library of Congress* (pp. 49). Cataloging Policy and Support Office.

Library OPACs containing UDC codes. Retrieved March 21, 2017, from http://www.udcc.org/index.php/site/page?view=opacs

Linked data. from https://www.w3.org/standards/semanticweb/data

Losee, R. M. (1993). Seven fundamental questions for the science of library classification. *Knowledge Organization, 20*, 65-65.

Maddaford, S., & Briefing, C. Library of Congress Classification System.

Madge, O.-L. (2011). Evidence Based Library and Information Practice. *Studii de Biblioteconomie şi Ştiinţa Informării*(15), 107-112.

Mai, J.-E. (2003). The future of general classification. *Cataloging & classification quarterly, 37*(1-2), 3-12.

Mai, J.-E. (2004). Classification in context: relativity, reality, and representation. *Knowledge Organization, 31*(1), 39-48.

Mai, J.-E. (2005). Analysis in indexing: document and domain centered approaches. *Information Processing & Management, 41*(3), 599-611. doi: http://dx.doi.org/10.1016/j.ipm.2003.12.004

Mai, J.-E. (2009). The boundaries of classification.

Mai, J.-E. (2010). Classification in a social world: bias and trust. *Journal of Documentation, 66*(5), 627-642.

Mai, J.-E. (2011). The modernity of classification. *Journal of Documentation, 67*(4), 710-730.

Mai, J. (2004). Classification of the Web: challenges and inquiries. *Knowledge Organization, 31*(2), 92.

Mancuso, J. (1994). *General Purpose vs Special Purpose Couplings.* Paper presented at the 23rd Turbomachinery Symposium, Dallas, TX, Sept.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1): Cambridge university press Cambridge.

Maron, M. E., Kuhns, J. L., & Ray, L. C. (1959). *Probabilistic indexing: a statistical approach to the library problem*. Paper presented at the Preprints of papers presented at the 14th national meeting of the Association for Computing Machinery, Cambridge, Massachusetts.

McGrath, K. (2007). Facet-based search and navigation with LCSH: Problems and opportunities. *code4lib Journal, 1*.

McIlwaine, I. C. (1997). The Universal Decimal Classification: Some factors concerning its origins, development, and influence. *Journal of the American Society for Information Science (1986-1998), 48*(4), 331.

Miksa, S. D. (2007). The challenges of change: a review of cataloging and classification literature, 2003-2004. *Library resources & technical services, 51*(1), 51.

Neelameghan, A., & Lalitha, S. (2013). Multilingual thesaurus and interoperability. *DESIDOC Journal of Library & Information Technology, 33*(4).

Neelameghan, A., & Parthasarathy, S. (1997). *SR Ranganathan's Postulates and Normative Principles: Applications in Specialized Databases Design, Indexing and Retrieval*: Sarada Ranganathan Endowment for Library Science.

Nizamani, S., Memon, N., & Wiil, U. K. (2011). Cluster Based Text Classification Model *Counterterrorism and Open Source Intelligence* (pp. 265-283): Springer.

Painter, A. F. (1974). Classification: Theory and Practice. *Drexel Library Quarterly, 10*(4), n4.

Panigrahi, P., & Prasad, A. (2005). Inference Engine for Devices of Colon Classification in AI-based Automated Classification System.

Perles, B. (1995). Faceted Classifications and Thesauri. Retrieved from Howard Besser's Web website: http://besser.tsoa.nyu.edu/impact/f95/Papers-projects/Papers/perles.html

Peroni, S. (2012). *Semantic Publishing: issues, solutions and new trends in scholarly publishing within the Semantic Web era.* alma.

Piros, A. (2014, 29 February 1, 2014). *A different approach to Universal Decimal*

Classification in a Mechanized Retrieval

System. Paper presented at the Proceedings of the 9th International Conference on Applied Informatics

Eger, Hungary.

Pollitt, A. S. (1998). The key role of classification and indexing in view-based searching: Technical report, University of Huddersfield, UK, 1998. http://www.ifla.org/IV/ifla63/63polst.pdf.

Press, O. F. (2002). Introduction to the dewey decimal classification.

Raghavan, K. (2016). The Colon Classification: A few considerations on its future. *Annals of Library and Information Studies (ALIS), 62*(4), 231-238.

Rahman, A., & Ranganathan, T. (1962). Seminal Mnemonics. *Annals of Library Science, 9*, 53-67.

Ramesh, P., Vivekavardhan, J., & Bharathi, K. (2015). Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. *DESIDOC Journal of Library & Information Technology, 35*(3).

Ranganathan, S. R. (1968). Choice of scheme for classification. *Lib. Sci. with a slant to Documentation, 5*(1), 1-69.

Reiner, U. (2008). Automatic analysis of Dewey decimal classification notations *Data analysis, machine learning and applications* (pp. 697-704): Springer.

Rodriguez, R. D. (1984). Hulme's concept of literary warrant. *Cataloging & classification quarterly, 5*(1), 17-26.

Rosenfeld, L., & Morville, P. (2002). *Information architecture for the world wide web*: " O'Reilly Media, Inc.".

San Segundo Manuel, R. (2008). Some arguments against the suitability of Library of Congress Classification for Spanish Libraries. *Extensions and Corrections to the UDC*.

Sateli, B., & Witte, R. (2015). *Automatic construction of a semantic knowledge base from CEUR workshop proceedings.* Paper presented at the Semantic Web Evaluation Challenge.

Satija, M. P. (2013). *The theory and practice of the Dewey decimal classification system*: Elsevier.

Satija, M. P. (2015). Save the national heritage: Revise the Colon Classification.

Satija, M. P., & Martínez-Ávila, D. (2015). Features, Functions and Components of a Library Classification System in the LIS tradition for the e-Environment. *Journal of Information Science Theory and Practice, 3*(4), 62-77.

Satija, M. P., & Singh, J. (2010). Colon Classification (CC) *Encyclopedia of library and information sciences* (Vol. 2, pp. 1158-1168).

Schallier, W. (2005). *Subject retrieval in OPAC's: a study of three interfaces*. Paper presented at the 7th ISKO-Spain Conference: The human dimension of knowledge Organization, Barcelona.

Singapore, N. L. o. (2016). Usability on the Web. from http://www.nlb.gov.sg/resourceguides/usability-on-the-web/

Slavic-Overfield, A. (2005). *Classification management and use in a networked environment: the case of the Universal Decimal Classification.* University of London.

Slavic, A. (2006). Interface to classification: some objectives and options.

Slavic, A. (2008). Use of the Universal Decimal Classification: A world-wide survey. *Journal of Documentation, 64*(2), 211-228.

Smiraglia, R. P., & Van den Heuvel, C. (2011). Idea Collider: From a theory of knowledge organization to a theory of knowledge interaction. *Bulletin of the American Society for Information Science and Technology, 37*(4), 43-47.

Subject classification schemes. (2015). from http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/9042

Sukhmaneva, E. (1970). The Problems of Notation and Faceted Classification. *17*(3-4), 112-116.

Svenonius, E. (2000). *The intellectual foundation of information organization*: MIT press.

Tallerås, K. (2013). From Many Records to One Graph: Heterogeneity Conflicts in the Linked Data Restructuring Cycle. *Information Research: An International Electronic Journal, 18*(3), n3.

Tennis, J. T. (2008). Epistemology, theory, and methodology in knowledge organization: toward a classification, metatheory, and research framework.

Tennis, J. T. (2011). Ranganathan's layers of classification theory and the FASDA model of classification.

Thelwall, M. (2009). Synthesis lectures on information concepts, retrieval, and services.". *Introduction to webometrics: Quantitative web research for the social sciences*.

Tomren, H. (2003). Classification, bias, and American Indian materials. *Unpublished work, San Jose State University, San Jose, California*.

Tunkelang, D. (2009). Faceted search. *Synthesis lectures on information concepts, retrieval, and services, 1*(1), 1-80.

Ullah, A., Ullah, I., Khusro, S., & Ali, S. (2016, 19-21 Dec. 2016). *BookOnt: A Comprehensive Book Structural Ontology for Book Search and Retrieval.* Paper presented at the 2016 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan.

Ullah, I., & Khusro, S. (2016). In Search of a Semantic Book Search Engine on the Web: Are We There Yet? *Artificial Intelligence Perspectives in Intelligent Systems* (pp. 347-357): Springer.

Universal Decimal Classification summary. (2017). from http://www.udcsummary.info/php/index.php?id=67277&lang=en#

Vizine-Goetz, J. S. M. D. (2009). The Dewey Decimal Classification. *Encyclopedia of Library and Information Science*.

Wang, J. (2009). An extensive study on automated Dewey Decimal Classification. *Journal of the American Society for Information Science and Technology, 60*(11), 2269-2286.

Wijewickrema, C. M., & Gamage, R. (2013). An ontology based fully automatic document classification system using an existing semi-automatic system.

Xin, R. S., Hassanzadeh, O., Fritz, C., Sohrabi, S., & Miller, R. J. (2013). Publishing bibliographic data on the Semantic Web using BibBase. *Semantic Web, 4*(1), 15-22.

Yelton, A. (2011). A Simple Scheme for Book Classification Using Wikipedia. *Information Technology and Libraries, 30*(1), 7-15.

Yi, K. (2006). *Challenges in automated classification using library classification schemes.* Paper presented at the Proceedings of world library and information congress: 72nd ifla general conference and council.

Zhu, Z. (2011). *Improving Search Engines via Classification.* University of London.

doi:10.6017/ital.v36i3.8930