

Analysis of mapping types for terminology services

Emma McCulloch and George Macgregor

Centre for Digital Library Research, Department of Computer & Information Sciences, University of Strathclyde, Glasgow, UK

*Correspondence to: Emma McCulloch, Centre for Digital Library Research, Department of Computer & Information Sciences, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH. UK.
E-mail: e.mcculloch@strath.ac.uk*

Abstract

This paper assesses the range of mapping types required to facilitate interoperability in the context of a distributed terminology server. A detailed set of match types were examined, with a view to determining their validity for characterising relationships between mappings from selected terminologies (AAT, MeSH, LCSH and UNESCO) to the Dewey Decimal Classification (DDC) scheme. It was hypothesised that the detailed set of 19 match types proposed by Chaplan (1995) is unnecessary in this context and that they could be reduced to a less detailed conceptually-based set. Results from an extensive mapping exercise support the main hypothesis and a generic suite of match types are proposed, although doubt remains over the current adequacy of the developing Simple Knowledge Organization System (SKOS) Core Mapping Vocabulary Specification (MVS) for inter-terminology mapping.

Keywords: Classification; Interoperability; Knowledge Organization Systems; SKOS Core; Terminologies; Vocabulary mapping

1. Introduction

The recent growth in distributed digital libraries and repositories has restored interest in the interoperability of Knowledge Organization Systems (KOS) to facilitate user access to discrete heterogeneous digital objects [1]. KOS employ a variety of disparate *terminologies* in the form of term lists (e.g. authority files, glossaries, gazetteers, dictionaries), classifications and categorisation schemes (e.g. bibliographic classifications, taxonomies, categorisation schemes) and relational vocabularies (e.g. thesauri, subject heading lists, semantic networks, ontologies) [2].

Within the growing number of repositories, digital objects will require to be indexed and organised in accordance with a variety of different schemes. Since it is unrealistic to expect users to interrogate each repository separately or to familiarise themselves with the numerous terminologies deployed, it is increasingly important that users are able to search or browse multiple distributed repositories simultaneously. Currently, however, the accuracy of such systems tends to remain dependent upon the degree of interoperability afforded between the terminologies in use. Technical approaches seeking to artificially or intellectually optimise interoperability therefore continue to form a key area of research [e.g. 2, 3, 4, 5]. One such approach that has attracted significant attention is *terminology mapping* (or vocabulary mapping).

Terminology mapping is evident in a variety of KOS interoperability approaches and essentially involves imposing equivalence, conceptual and hierarchical relationships between terms in different schemes [4]. The assumption underpinning mapping is that equivalence can exist between disparate KOS and their respective terminologies [4]; however, exact equivalence is rarely attainable [6]. Whilst recent research into the application of automated techniques has aided in the management of large terminology sets and even assisted in mapping implementation itself [7], the process of terminology mapping remains largely intellectual, and therefore heavily dependent on human intervention. One continuing problem inherent in the terminology mapping process - whether intellectual or automated - is accurately characterising the type of mapping match between terms. The existence of linguistic inconsistencies across terminologies (e.g. synonyms, homonyms, antonyms, etc.), grammatical variations (e.g. singular / plural forms, alternative spellings or punctuation, verb tenses, etc.), variations in subject coverage, and the relative specificity or level of granularity with which terminologies accommodate like concepts, limit their potential for exact equivalence. Differing semantic structures of the terminologies being mapped can also prove problematic for mapping across different KOS; for example, classification schemes have a radically different structure to relational vocabularies. Consequently, mapped terms may only exemplify partial equivalence.

Given that exact equivalence between terminologies will be rare, it is necessary to accurately characterise the degree of equivalence by assigning match types during the mapping process. This is considered necessary to:

- Enable the ranking of results according to the degree of concordance with users' preferred terminology.
- Provide users with details of the precise nature of the relationship(s) between their entered query and their retrieved result set (which will invariably include mapped terms from other terminologies, or comprise resources retrieved using terms derived from mapped terminologies).
- Impart sufficient information during subject hierarchy browsing to enable users to make informed decisions about the relevance of mapped terms.
- Provide users with mappings that can be used to generate relevance feedback.
- Help identify mapping regularities between specific terminologies, thus facilitating the research and development of improved automated routines to assist in large-scale terminology mapping.

Various match types have been proposed, e.g. [3, 4, 7, 8, 9, 10]. In this paper we examine terminology mapping match types in relation to a Dewey Decimal Classification (DDC) based terminology server. In particular, we assess the suitability of Chaplan's 19 match types [8] as forming the basis of a *generic suite* of equivalence matches to be used by services employing terminology mapping.

The remainder of this paper is organised as follows: sections 2, 3 and 4 will review related literature and establish the aims of the study. Section 5 describes the methodology used to test the Chaplan match types in relation to our data set. The crux of the paper (sections 6 and 7) deals with the results, analysis and subsequent discussion. Conclusions and suggestions for further research are provided in section 8.

2. Terminology mapping

Interest in mapping as a means of facilitating terminology interoperability for improved distributed searching is not new. The rapid development of distributed online databases in the 1960s and 1970s, and the associated rise in domain-specific terminologies, forced researchers to address the issue of terminology compatibility and related system-based solutions, with terminology mapping featuring significantly in many of the proposed solutions. Although so-called *direct mapping* was popular and continues to be for some recent solutions [11, 12], it generally requires considerable intellectual effort and resourcing [2, 13]. Mapping work has consequently focussed on the use of terminology *switching* (or *switching languages*) to simplify the management of multiple terminological mappings and to minimise the intellectual demands normally associated with direct mapping.

The switching model entails the use of a single terminology as an intermediary (Figure 1). Each of the terminologies to be used in the retrieval system (A – H) is mapped to a common terminology (X). This allows user queries entered using terminology A to be translated to X and then switched to the equivalent terms in terminology G, for example. Switching was often the chosen model for early mapping research [14, 15, 16] and has recently been revisited [3, 17].

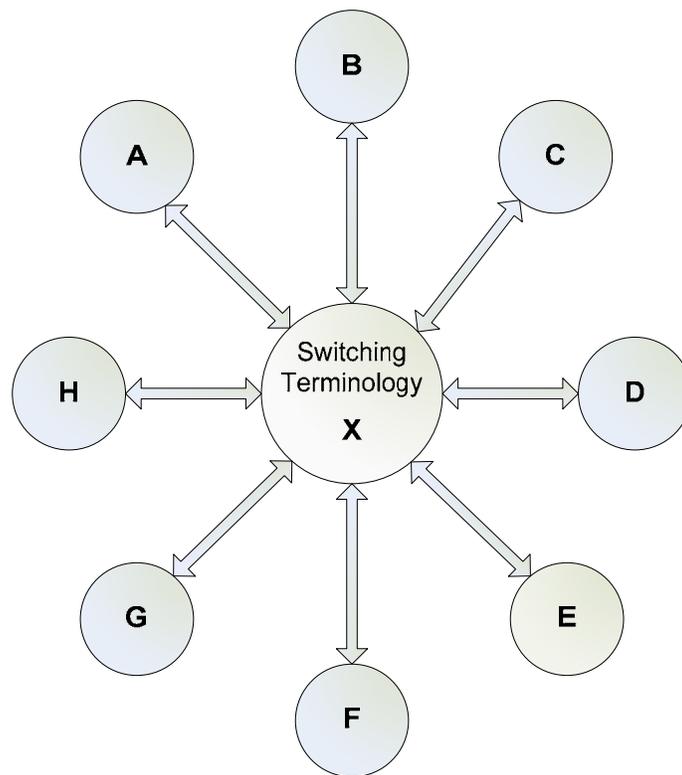


Figure 1. Typical terminology switching model

Any terminology can theoretically be used for switching although it is generally acknowledged that the coverage of the selected switching terminology must be sufficiently broad to include most, if not all, disciplines encompassed by the schemes with which it is to be used [15, 18]. Failure to select such a broad terminology will result in the switching language degrading any requests sent from a detailed terminology (e.g. a domain-specific terminology). The use of universal classification schemes for switching has therefore attracted much attention, particularly those schemes that are decimal-based. The theoretical foundation for using such terminologies was established early on [19, 20] and has subsequently been explored by a variety of recent research projects, e.g. [3, 21, 22]. Schemes such as the Universal Decimal Classification (UDC) and the Dewey Decimal Classification

(DDC) exemplify wide subject coverage and are suitable for multi-disciplinary user groups. They also experience global use; DDC, in particular, is available via the subscription-based OCLC Connexion service, providing quick access to frequently updated schedules and thus aiding the mapping and management of terminologies [3]. The analytico-synthetic features, although more prominent in UDC than DDC, allow the construction of diverse and detailed concepts which can then be expressed in hierarchical notation [23]. More generally, such schemes are conducive to hierarchical browsing and facilitate the display of associated thesaurus terms [2].

The match type work documented in this paper is based within the context of a Dewey Decimal Classification (DDC) spine-based terminology server [24]. The spine-based approach is conceptually similar to switching but differs in that DDC does not always assume a passive role in the process. Rather, DDC remains central to users' *disambiguation* processes [25]. Since DDC notation is generally indicative of the taxonomic hierarchy, the truncation of DDC numbers is initiated if no hits are found using mapped terminologies. This truncation occurs successively until one or more hits are identified. For example, if a user query for 'Greenhouse gases' was matched to 363.73874 and no hits were found, the system will truncate the DDC number as follows:

- 363.7387 (Fumes, gases, smoke)
- 363.738 (Pollutants)
- 363.73 (Pollution)

The spine-based approach also facilitates hierarchical browsing and the discovery of like terms within other terminologies. The current Machine to Machine (M2M) web-service implementation of this server provides a variety of terminological functions, such as the enrichment of users' search queries by providing (where applicable) related terms (RT), broader terms (BT), narrower terms (NT), scope notes, etc. associated with specific, named, terminologies (which are then marked up in the Simple Knowledge Organization System (SKOS) Core and sent to local systems for use) [26]. However, the primary function of the server remains mapping between disparate terminologies and within a variety of different user searching scenarios.

3. Match types: related work

Several investigations into mapping match types have been conducted over many years and have largely arisen as a result of research into terminology mapping; however, investigations into terminology compatibility and integration have also been successful in defining degrees of equivalence. For example, Neville [9] studied the types of incompatibility between keywords in different thesauri pertaining to the same subject area for possible thesauri reconciliation based on a source thesaurus using concept code numbering. Although he identified

numerous types of relationships and proposed some solutions for accommodating those which were complex, many have limited applicability within an operational mapping system.

While researching and developing an operational terminology switching system, Silvester and Klingbiel [10] developed a series of rules to accommodate switching between the Defense Technical Information Center (DTIC) subject terms and the National Aeronautics and Space Administration (NASA) thesaurus by using a so-called 'lexical dictionary'. These rules were established as system commands, but characterised the degree of term equivalence between terminologies: with delete, identity, simple change, list, and Table. 'Delete' indicates that there are no conceptually equivalent terms in the NASA thesaurus (i.e. no match between input and output terms). 'Identity' indicates that the input is identical or equal to the output (i.e. an exact match between input and output terms). 'Simple change' indicates a 'minor change' and characterises instances where the input term expresses the same concept as the output but differs in minor respects. For example, the input term may be plural and the output term is singular. This rule also accommodates other grammatical variations such as suffix variations (e.g. 'ing' or 'tion') and synonyms. 'List' applies when a single term is switched to multiple terms. In such instances the multiple terms express the same concept (e.g. Machmeters / Mach number, speed indicators). Finally, 'Table' indicates the occurrence of 'Tables'. That is, the input term is context sensitive and requires additional terms to clarify the concept.

3.1. *Conceptual approaches: Renardus, SKOS Core, the Semantic Web*

Koch *et al.* [3] developed a Web-based service (Renardus) to facilitate searching and browsing across a variety of distributed European information services and subject gateways. DDC was used as a common switching terminology and browsing structure. Koch *et al.* acknowledged the need to specify the degree of mapping equivalence and used the principle of set theory to create five separate mapping match types: fully equivalent, narrower, broader, major overlap and minor overlap, when compared with a DDC class. It is worth noting that the Renardus match types are less concerned with expressing the specific nature of matches (or otherwise) and instead seek to characterise relationships of a *conceptual* nature. For example, 'fully equivalent' denotes that there is good equivalence between the terminologies, irrespective of how that concept may be represented. Such a match type essentially subsumes those matches generally described as exactly or conceptually equivalent [8, 10]. It also subsumes those matches that might normally be differentiated on the grounds of grammatical or lexical variations (e.g. plural/singular, abbreviations/acronyms, etc.). Whilst the approach proposed by Koch *et al.* [3] jettisons the emphasis placed on terminological incongruities, it is consistent with traditional classification and indexing theory which attempts to reconcile concepts rather than the terms used to represent those concepts.

A similar approach has been adopted by semantic web technologies, such as the proposed W3C Simple Knowledge Organization System (SKOS) Core [27]. SKOS Core is based on the representation of concepts and is an application of the Resource Description Framework (RDF). It provides a model for expressing the structure and content of various KOS to enable easy machine processing. Miles and Brickley [28] have proposed the SKOS Core Mapping Vocabulary Specification (MVS) to support the mapping of concepts between different schemes using the SKOS Core framework. This emerged from similar work [29] undertaken by the SWAD-Europe project [30]. The properties proposed by SKOS are: *exactMatch*, *broadMatch*, *narrowMatch*, *majorMatch* and *minorMatch*. The SKOS Core MVS also supplements the match types with a series of classes (AND, OR, NOT) for combining or excluding concepts. For example, the class AND is used to denote the intersection of two or more concepts. The term of *Health services administration* in terminology A may therefore map to *Health services AND administration* in terminology B.

The definitions of the SKOS MVS match types are similar to those used by Koch *et al.* [3] and are based on the assumption that the number of resources assigned to a particular concept is known. For example, *majorMatch* is where a "set of resources properly indexed against concept A share more than 50% of its members with the set of resources properly indexed against concept B" [28]. It therefore remains unclear how appropriate the SKOS Core MVS currently is for terminology mapping services. Such match type definitions are conducive to static terminology mappings, but less suited to dynamic mappings (invoked via a terminology server) where little is known about the resources or the indexes held in the repositories with which a client will interact. Although some of the match types could theoretically be used, their application would probably be inconsistent with the conceptual underpinnings and assumptions inherent in the Specification (unless appropriate extensions or modifications are made). The SKOS Core MVS has yet to experience wide deployment or testing; however, Liang *et al.* [31] report difficulties while mapping from AGROVOC Thesaurus to the Chinese Agricultural Thesaurus. They cite ill-defined mapping properties and find the assumptions inherent in the Specification to limit particular applications. Liang *et al.* consequently propose some redefinitions.

3.2. *Matches derived via co-occurrence mapping*

OCLC have experimented significantly with *co-occurrence mapping*, e.g. [32, 33] involving statistical routines which extract 'loosely-mapped' terms from metadata records containing terms from more than one terminology [2]. For example, it is possible to implement a co-occurrence process using MARC21 Authority Format [34] metadata records that employ tag 082 (DDC number) and tags 600-651 (subject added entry) that use second indicator 0 (denoting LCSH) in order to derive a loose set of mappings between the two terminologies. Such techniques have been used to great effect by OCLC to provide popular Library of Congress Subject Headings (LCSH) with mapped DDC numbers for practitioners [35] and within the WebDewey service [36].

Vizine-Goetz *et al.* [7] recently conducted research to further develop such inter-vocabulary association techniques and mapped terms from the Educational Resources Information Center (ERIC) Thesaurus to LCSH. Their methodology entailed encoding both terminologies according to the MARC21 Authority Format and implementing a series of algorithms to ascertain matches. To aid match analysis and to express inter-term relationships, Vizine-Goetz *et al.* categorised matches according to four separate match types: PT/PT, PT/NPT, NPT/NPT and NPT/PT all of which signify exact matches between preferred and non preferred terms in source/target terminologies. While spacing, capitalisation and punctuation are ignored during their matching process, Vizine-Goetz *et al.* acknowledge that they focus on exact matches and that various other potential matches (e.g. plural/singular, further specification, etc.) are not accommodated within their categorisation. The match types can not therefore be said to be exhaustive and are optimised for investigation of the relational vocabularies at hand (i.e. ERIC and LCSH). Their use as generic mapping types (i.e. applicable to all kinds of KOS) consequently remains unclear.

3.3. Chaplan match types

Arguably the most significant contribution to mapping match types has been proposed by Chaplan [8] whose investigation focussed on identifying the nature of term matches that could potentially be used to enhance the performance of switching systems. Chaplan's methodology entailed the intellectual mapping of terms from the Laborline Thesaurus to LCSH, resulting in the subsequent identification of 19 separate match types (Table 1). Chaplan concluded that the relationships between terminologies were 'vastly more complex than supposed' and stated that simple conceptual matches (i.e. exact match, partial match, no match) were inadequate to accurately characterise the full range of relationship types evident between terms in different schemes. It is note worthy that several of Chaplan's more complex matches confirm those identified during thesauri reconciliation experiments by Neville [9] (e.g. Chaplan: Opposite or negative; Neville: Antonymous terms). Chaplan notes that further research is required to ascertain whether these results are applicable across a variety of different terminologies. The work documented in this paper goes some way to testing the applicability of Chaplan's match types across a variety of KOS, details of which are provided in the methodology section.

Match type code	Definition	Chaplan's examples Laborline Thesaurus	Chaplan's examples LCSH
1	Exact match	Industrial relations	Industrial relations
2	Exact cross-reference match	Child labor	USE Children—employment

3	Exact match, but with intervening characters	Research management	Research—management
4	Plurals	Displaced worker	Displaced workers
5	Subordination, in the form of a species-genus relationship	Industrywide bargaining	Collective bargaining
6	Superordination, in the form of genus-species relationship	Motor vehicle industry	Automobile industry and trade
7	Part-of-speech difference	Employment interview	Employment interviewing
8	Word-order variation	Illegal alien	Aliens, illegal
9	Further specification	Absenteeism	Absenteeism (labor)
10	Spelling variation	No strike clause	No-strike clause
11	Suffix variation	Quality of working life	Quality of work life
12	Abbreviation or acronym	Alta.	Alberta
13	Subdivision (Represents term that was used only as a subdivision in LCSH)	Measurement	Measurement
14	Concept match	Performance appraisal	Employees—Rating of
15	Homograph	Millinery [referring to hat industry]	Millinery [referring to costume hats]
16	Translation	Precedent	Stare decisis
17	Date or numerical variation	1935	Nineteen thirty-five
18	No match	Boulwarism	deskilling
19	Opposite or negative	Desegregation	Segregation

Table 1. Chaplan's terminology match types.

4. Rationale and Objectives

The work documented here attempts to examine terminology mapping match types in relation to a DDC-based terminology server. In particular, we assess the suitability and validity of Chaplan's 19 match types [8] as the basis of a *generic suite* of equivalence matches to be used by services employing terminology mapping. Such an assessment requires consideration of whether mappings between disparate KOS terminologies can be adequately represented by Chaplan's set of match types or whether alternative and/or additional match types are required.

This is particularly important since the majority of match type research focuses on mapping between similarly structured relational vocabularies.

An earlier instantiation of the DDC spine-based terminology server used match types based on the work of Chaplan. These match types were used primarily to aid the user during the disambiguation process whereby they select their preferred term in context from the DDC hierarchies presented. The growth of semantic web applications [37] and the associated need to deconstruct and link lexically disparate search terms or phrases [38] suggests that the broad range of match types proposed by Chaplan may not always be required. However, the conceptual approach based on set theory (i.e. SKOS Core MVS), as noted in section 3.1, is currently limited for terminology services and we envisage instances within our framework where finer granularity may be required (e.g. during particular phases of user disambiguation). Similar to Liang and Sini [39] we consider the conceptual approach to be somewhat abstract for the practical application of mappings in this context. Extensions to the SKOS-Core MVS are outside the scope of this paper; we are interested in the extent to which Chaplan's match types could form the basis of a generic suite of match types to be used by terminology services. It is hypothesised that such a large number of match types - across a variety of terminologies and using the specified rules - is unnecessary and could easily be collapsed into smaller number, possibly reflecting alternative approaches [3, 28]. It is also thought that the scope of some match types (specifically part of speech difference and suffix variation) are ill-defined, which may lead to misapplication.

5. Methodology

5.1. Selection of schemes

To test the validity of Chaplan's match types, four terminologies were selected for mapping to DDC [40]: LCSH [41], MeSH (Medical Subject Headings) [42], UNESCO Thesaurus [43] and AAT (Art and Architecture Thesaurus) [44]. The selection of these particular terminologies was purposive. Each of the selected terminologies experiences wide international use and two (MeSH and AAT) are discipline specific, thus exemplifying significant subject detail and higher levels of granularity than the two general schemes (LCSH and UNESCO Thesaurus). A categorisation of the terminologies used according to Zeng and Chan's [2] KOS typology is provided in Table 2. An example of term lists was not included in the investigation; however, our assumption is that term lists exemplify simpler structures than relational vocabularies. Any match types capable of accommodating the latter form of KOS should theoretically be more than capable of accommodating the former.

Terminology	KOS Type
AAT	Relational vocabulary (thesaurus)
DDC	Classification and categorisation scheme (bibliographic classification scheme)
LCSH	Relational vocabulary (subject heading list)
MeSH	Relational vocabulary (thesaurus)
UNESCO Thesaurus	Relational vocabulary (thesaurus)

Table 2. Terminologies categorised using Zeng and Chan's KOS typology.

5.2. Selection of terms

Machine readable copies (in XML) of the terminologies were obtained and loaded into an appropriately structured database. A simple Java program was written to randomly select 50 terms from each terminology. The extracted terms were then mapped to DDC by both authors (A and B). To assist in the mapping process, terms from the selected terminologies and the terminological spine (DDC) were considered in context. That is, the nomenclature surrounding terms (in both extracted terms and DDC), any broader and narrower relationships, related terms, scope notes associated with terms, were all studied to ensure accuracy of mappings between terminologies. The nearest broader or narrower term was considered if no suitable exact or concept match could be found in the target terminology [4]. WebDewey [36] was used to search and browse DDC schedules for appropriate mappings. These tasks were undertaken *independently* by each author in order to increase the validity of identified mappings, and results were recorded in an appropriately structured matrix (Table 3).

Further consistency was ensured by observing strict DDC application rules with respect to the 'class here' and 'including' notes, which were treated distinctly [45]. For example, DDC caption scope notes employing the use of 'class here' are considered to *approximate the whole class* and therefore are unlikely to receive separate numbers. When instructed to 'class here' a concept match was assumed. Similarly, 'include' notes were considered to constitute a narrower term match. As such, where 'class here' and 'including' notes were evident, between-term relationships were coded as match types 14 and 5 respectively. Authors A and B re-grouped following the mapping process to compare results. Contentious mappings were examined closely and resolved through a process of reanalysis of DDC schedules and any available instructions relating to the mapped schemes.

UNESCO term	DDC no.	DDC captions	Auxiliary notation used	Optional notes

Vocational schools	373.246	Secondary education > Secondary schools and programs of specific kinds, levels, curricula, focus > Academic, military, vocational schools > Vocational schools	N/A	
Fuel technology	662.6	Chemical engineering > Technology of explosives, fuels, related products > Fuels	N/A	
Aquaculture	639.8	Agriculture > Hunting, fishing, conservation, related technologies > Aquaculture	N/A	
Library technicians	023.3	Library & information sciences > Personnel management (Human resource management) > Technician positions	N/A	<i>UNESCO term within DDC scope notes.</i>
Paramedical personnel	610.690233	Medicine and health > Organizations, management, professions > Medical personnel and relationships > Allied health personnel	Notation added from elsewhere in schedules.	

Table 3. Portion of example UNESCO to DDC mapping matrix.

5.3. Categorisation of mappings

Authors A and B then categorised the mappings in accordance with Chaplan's 19 match types. These categorisations were undertaken independently and were encoded by adding 1-19 to an additional column of the matrix. The authors then reconvened to determine the level of agreement of codes assigned across all 200 mapped terms. Individual matrices were merged to ascertain where authors' match types agreed, or otherwise. Where concordance on match types did not occur, the relevant terminologies were revisited to clarify terms in context together with relevant nomenclature. Where necessary, additional research work was undertaken, including consulting reference works and domain-specific resources to elucidate term definitions and scope. In all instances, the authors were able to reach agreement on match codes assigned.

5.4. Caveat

Chaplan's study assigned multiple match codes to mappings. This practice was not followed in the present study since the authors neither fully understood nor agreed with Chaplan's documented work. For example, the relationship between the terms 'watch making' and 'clock and watch making' could – according to Chaplan's

definitions – be simultaneously considered as 2 (exact cross-reference match), 10 (spelling variation), 15 (homograph) and 18 (no match). In the authors’ opinion a subordinate/superordinate relationship is also valid in such an example, but it is unclear why Chaplan has not encoded it accordingly. In addition, examples given to illustrate some match codes are ambiguous. For example, match code 8 (word order variation) is characterised as follows: A: Illegal alien; B: Aliens, illegal. Since this mapping also constitutes a singular/plural relationship it is not considered exclusive and is therefore a poor example with which to define word order variation. As a result of such uncertainty, and with a view to providing clarity to the user, the methodology asserted that only one match code could be assigned to any given mapping.

6. Findings

6.1. Match codes: level of agreement

The two sets of emergent data were combined to determine any areas of disagreement regarding the match codes assigned. The mean level of agreement between authors across all schemes was 164 (82%) with a standard deviation of 13.54. It was found (Table 4) that the level of agreement between authors was higher for discipline-specific schemes such as AAT and MeSH and somewhat lower for more generic schemes like LCSH and UNESCO. Taken together, the mean level of agreement for discipline-specific schemes was 93%, compared with 71% for the two general schemes investigated.

	AAT-DDC	LCSH-DDC	MeSH-DDC	UNESCO-DDC
Level of agreement	88%	74%	98%	68%

Table 4. Level of agreement between match codes assigned by authors A and B.

Authors A and B did not agree on the match type relationships between mapped terms on 36 of 200 occasions (18%) (Table 5). The highest proportion of disagreement was found between match codes assigned for LCSH and UNESCO, when compared to AAT and MeSH. The former two schemes elicited 80.56% of all disagreements between assigned match codes.

Match codes assigned by authors A/B or B/A	AAT-DDC	LCSH-DDC	MeSH-DDC	UNESCO-DDC	TOTAL
18/5	-	1	-	-	1

14/1	1		-	-	1
14/5	5	7	1	12	25
14/6	-	2	-	2	4
14/9	-	3	-	-	3
11/7	-	-	-	1	1
1/10	-	-	-	1	1
TOTAL	6	13	1	16	36

Table 5. Instances of conflict between match codes assigned by authors A and B.

A total of 33 of the 36 (87.88%) between-author disagreements involved match code 14. Such disagreement constituted 91.67% of all disagreements across the 200 mappings implemented. That is, on 33 occasions one author categorised a mapping as a concept match while the other considered it to demonstrate an alternative type of equivalence. In 29 of the 33 cases, conflict arose between a concept match (14) and narrower (5) or broader (6) term matches. Disagreements involving match code 14 were the only type encountered when mapping AAT and MeSH to DDC.

A further three distinct mismatches were evident from the LCSH and MeSH data (see Table 5). The first arose between code 18 (no match) and 5 (species-genus subordination). The second between code 11 (suffix variation) and 7 (part-of-speech difference), and the third between code 1 (exact match) and 10 (spelling variation). The latter was the only instance where authors A and B encoded a mapping differently and subsequently agreed on a third code when conflating their data. In all other cases, agreed match codes were consistent with at least one of the author's original categorisations.

6.2. Agreed match codes: frequencies

A frequency count of each of the agreed match codes was conducted (Table 6 and Figure 2). When mapping terms from AAT, LCSH, MeSH and UNESCO to DDC, match codes 1, 5 and 14 proved valid across all schemes. That is, terms from all four schemes elicited relationships categorised as exact match, narrower term and concept match when mapped to DDC (see Figure 2). The most commonly assigned match code was 5 (narrower term) constituting 113 (56.5%) of the 200 codes assigned. Beyond the 89% of mappings categorised as narrower, concept or exact matches, the remaining 11% were collectively indicative of match codes 3 (exact match with intervening characters), 4 (plural form), 6 (genus-species superordination (or broader)), 7 (part of speech difference), 9 (further specification) and 10 (spelling variation). Match code 6 (broader term) was only assigned

on three occasions, and was used to characterise the relationship between terms from the more general schemes of LCSH and UNESCO. Match code 6 was not applied when mapping from subject-specific schemes.

Match Code Number	AAT		LCSH		MeSH		UNESCO		Total Match Types Assigned Across All Terminologies	
	#	%	#	%	#	%	#	%	#	%
1	4	8	6	12	5	10	11	22	26	13
2	-	-	-	-	-	-	-	-	0	0
3	-	-	-	-	1	2	1	2	2	1
4	-	-	1	2	-	-	-	-	1	0.5
5	39	78	22	44	38	76	14	28	113	56.5
6	-	-	1	2	-	-	2	4	3	1.5
7	-	-	-	-	-	-	1	2	1	0.5
8	-	-	-	-	-	-	-	-	0	0
9	-	-	8	16	1	2	5	10	14	7
10	-	-	-	-	-	-	1	2	1	0.5
11	-	-	-	-	-	-	-	-	0	0
12	-	-	-	-	-	-	-	-	0	0
13	-	-	-	-	-	-	-	-	0	0
14	7	14	12	24	5	10	15	30	39	19.5
15	-	-	-	-	-	-	-	-	0	0
16	-	-	-	-	-	-	-	-	0	0
17	-	-	-	-	-	-	-	-	0	0
18	-	-	-	-	-	-	-	-	0	0
19	-	-	-	-	-	-	-	-	0	0
Total Terms	50	100	50	100	50	100	50	100	200	100

Table 6. Frequency count (and percentage) of assigned match codes, for individual schemes, and totals.

Mapping Types Assigned

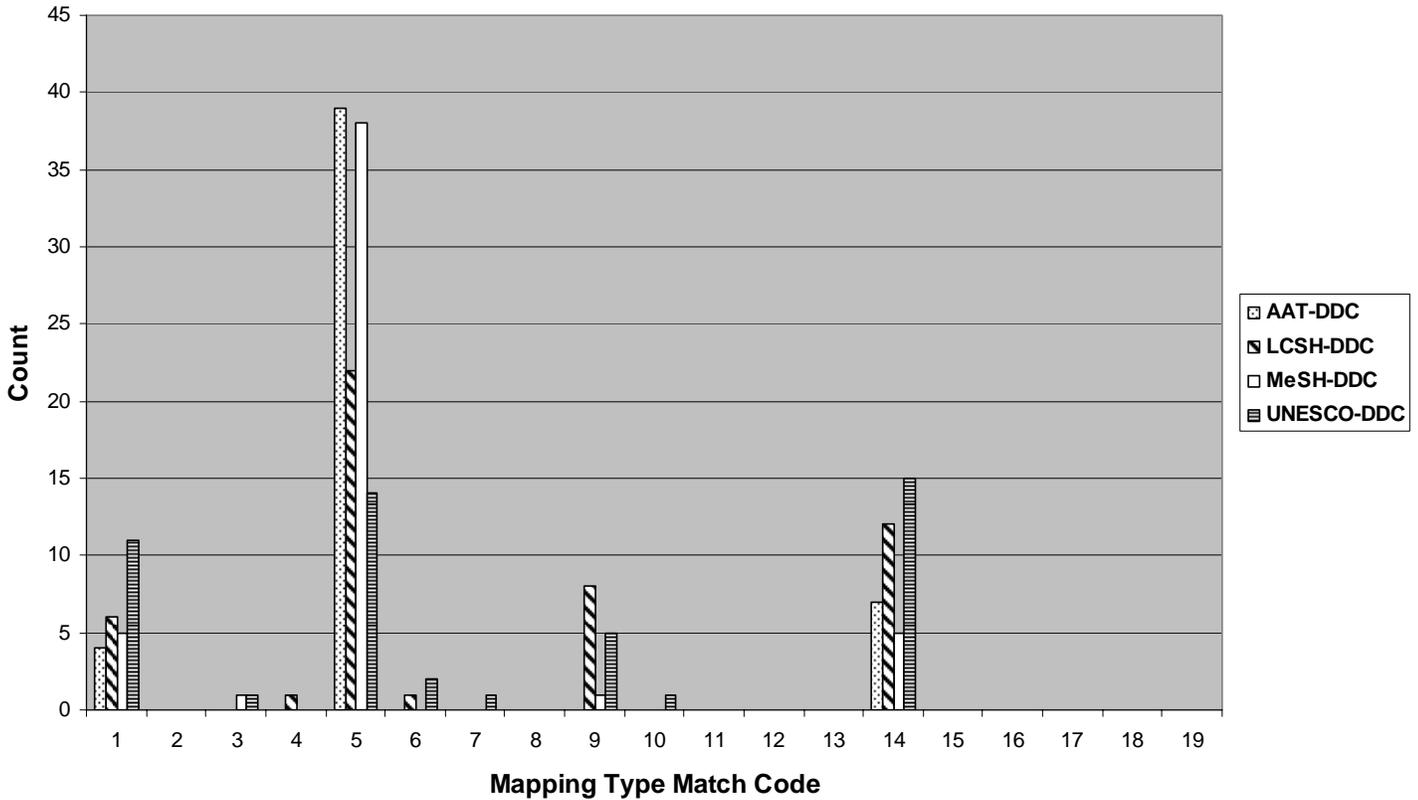


Figure 2. Frequency count of assigned match codes, by scheme.

6.3. Ranking of agreed match codes, by scheme

Assigned match codes were ranked according to frequency for each of the four schemes involved (see Table 7). Match code 5 (species-genus subordination) ranked the most frequently assigned code across three of the four schemes used - AAT, LCSH and MeSH - and ranked second in the case of UNESCO, with only a single occurrence (or 2%) separating the two top ranked match types in this case. UNESCO, when mapped to DDC, elicited one more concept match than narrower term match, making match code 14 the most highly ranked for this scheme.

The second most frequently assigned match code was 14: concept match. This was the case for AAT and LCSH. It also ranked joint second for terms mapped from MeSH to DDC along with code 1: exact match.

Exact matches (code 1) were the third most frequently assigned mapping type, constituting 13% of match types across all schemes. Exact matches were the third most frequently encountered relationship between terms mapped from AAT and UNESCO. MeSH mappings elicited an equal number of exact matches and concept matches (5 occurrences or 10%). For LCSH code 9 (further specification) ranked in third place.

Match code number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
AAT	3	-	-	-	1	-	-	-	-	-	-	-	-	2	-	-	-	-	-
LCSH	4	-	-	5.5	1	5.5	-	-	3	-	-	-	-	2	-	-	-	-	-
MeSH	2.5	-	4.5	-	1	-	-	-	4.5	-	-	-	-	2.5	-	-	-	-	-
UNESCO	3	-	7	-	2	5	7	-	4	7	-	-	-	1	-	-	-	-	-
Mean Ranking (to 2 d.p.)	3.13	0	5.75	5.5	1.25	5.25	7	0	3.83	7	0	0	0	1.88	0	0	0	0	0

Table 7. Ranking of assigned match codes, by scheme.

Table 8 summarises these findings by presenting assigned match codes by ranking, illustrating that fewer mapping types were required to characterise relationships between subject-specific schemes and DDC. A greater range of relationships was evident when considering equivalence relationships between terms in LCSH and UNESCO; that is to say, a more varied set of match codes was applied.

Ranking	AAT	LCSH	MeSH	UNESCO
1	5	5	5	14
2	14	14	1/14	5
3	1	9		1
4	-	1	3/9	9
5	-	4/6		6

6	-		-	3
7	-	-	-	7
8	-	-	-	10

Table 8. Assigned match codes by ranking and scheme.

6.4. Match types validated

Tables 7 and 8 indicate that a total of nine Chaplan match types were deemed valid for the purpose of expressing equivalence relationships from terms in AAT, LCSH, MeSH, and UNESCO, to DDC. Illustrative examples are provided in Table 9. Of the remaining ten relationship types identified by Chaplan only two were assigned throughout the study, and were subsequently ruled out following the merging of authors' encoded match types. On one occasion author A categorised a match as 11 (suffix variation); and on a second occasion as 18 (no match). In the former case, the assignment of code 11 was replaced with 7 (part-of-speech difference). In the latter, 18 was replaced with 5 (narrower term).

Match code number	Match type	Scheme Term	DDC Term/Hierarchy
1	Exact Match	AAT: Strasbourg The AAT scope note states: <i>“Refers to the style of faience produced at the Strasbourg pottery and porcelain factory in the 18th century. Widely imitated throughout Europe, the style features naturalistic floral decoration rendered in brightly colored enamel.”</i>	DDC: Ceramic arts > Earthenware and stoneware > Historical and geographic treatment > Europe Western Europe > France and Monaco > Champagne-Ardenne, Ile-de-France, Lorraine, Alsace > Alsace > Bas-Rhin department > Strasbourg
3	Exact Match with Intervening Characters	UNESCO: Viet Nam	Southeast Asia > Vietnam
4	Plural Form	LCSH: Eye	Specific physiological systems in animals, regional histology and physiology in animals > Nervous and sensory systems > Eyes

5	Species-Genus Subordination	MeSH: Chromosomes, Human, Pair 5	Life sciences; biology > Internal biological processes and structures > General internal processes common to all organisms > Biochemistry > Specific biochemicals and biochemical genetics > Biochemical genetics > Chromosomes
6	Genus-Species Superordination	LCSH: Cultural industries	Production > Secondary industries and services > Services and specific products > Documentary media, educational media, news media; journalism; publishing > Publishing
7	Part-of-Speech Difference	UNESCO: Heating	Physics > (<i>Specific forms of energy</i>) > Heat
9	Further Specification	LCSH: Managed care plans (Medical care)	Social welfare problems and services > Physical illness > Special topics of physical illness > Social aspects > Forms of assistance > Managed care plans
10	Spelling Variation	UNESCO: Educational programmes	Labor, social service, education, cultural law > Education > Finance > Educational programs
14	Concept Match	AAT: Scottish	Culture and institutions > Historical, geographic, persons treatment > Treatment by specific continents, countries, localities; extraterrestrial worlds > Europe Western Europe > British Isles > Scotland

Table 9. Examples of the nine match types verified.

6.5. Key findings: summary

- Match codes were assigned more consistently for subject specific schemes than for more general schemes.
- 91.67% of between-author disagreements (as shown in Table 5) involved match code 14.
- A total of nine of Chaplan’s original 19 match types were verified.
- Exact matches, concept matches and narrower term matches were the three most frequently assigned match codes, and were the only three to prove valid across all four schemes investigated. Between them, they accounted for 178 of 200 (89%) codes assigned.
- A narrower range of match codes was required to categorise relationship types when mapping terms from subject-specific schemes to DDC, compared to that of the general schemes LCSH and UNESCO.

7. Discussion

7.1. Match codes: level of agreement

It was noted in section 6 that between-author variation arose in relation to particular assigned match codes. 91.67% of the said variation involved match code 14 (concept match) suggesting that the nature of this equivalence relationship is poorly defined, resulting in blurring of boundaries with other match codes. The fact that 80.56% of variations involving code 14 also involved codes 5/6 suggests that there is general confusion over what may constitute a narrower/broader term match and a concept match. It seems that concepts are often considered equivalent when one is actually a super/sub-set of the other. The blurring of concept matches and narrower/broader term matches could result from an inability to distinguish sufficiently between an equivalent concept and super/sub concepts (i.e. X is part of Y) or it may be a symptom of limited subject knowledge on the part of the authors. Besides disagreement involving code 14, a total of three additional disagreements were encountered.

Although evident on a single occasion only, the conflict between codes 7 and 11 suggests that such subtle linguistic distinctions are not required for terminology mapping within this context. In this instance the UNESCO term 'Heating' was mapped to DDC 'Physics > (Specific forms of energy) > Heat', in accordance with Chaplan's definitions. Author B considered this to be an example of 'part-of-speech difference' in line with Chaplan's example:

- **A:** Employment interview
- **B:** Employment interviewing

In contrast, author A categorised this mapping as a 'suffix variation', conforming to Chaplan's example:

- **A:** Quality of working life
- **B:** Quality of work life

On revisiting Chaplan's definitions there is no significant difference between the two examples quoted above, suggesting that they are sufficiently equivalent and could be merged. Indeed, Chaplan herself implied that these two measures of equivalence may not be sufficiently distinct from one another following a 50% overlap in terms categorised with both codes during her study.

Authors A and B assigned codes 18 and 5 respectively to the following example:

- **LCSH:** Don Juan (Legendary character)
- **DDC:** Rhetoric and collections of literary texts from more than two literatures > Collections of literary texts from more than two literatures > Arts and literature dealing with specific themes and subjects > Humanity > Specific persons

When revisited by the authors it became evident that a scope note under DDC Table 3-C [36] provides instruction to ‘Include Don Juan’. It therefore follows that code 5 was agreed upon.

Where disagreement arose from one author assigning code 1 (exact match) and the other code 10 (spelling variation), the overall outcome was inconsistent with both authors. This was the only single occurrence of neither authors’ codes being assigned following re-analysis of the terms in context. The authors agreed that example:

- **A:** Viet Nam
- **B:** Southeast Asia > Vietnam

should be assigned match code 3 (exact match with intervening characters). Since each character is exactly matched it was agreed that capitalisation did not constitute a spelling variation as such, but that the space in case A constituted – in a machine readable sense - an ‘intervening character’. It could be argued that the above example also provides ‘further specification’; however, the decision was taken early on that DDC hierarchies would be taken into consideration since contextual detail was required to ascertain, for example, whether a DDC caption was broader or narrower than its equivalent in an alternative scheme.

7.2. *Agreed match codes: frequencies*

It is likely that the reason match code 5 was the most frequently encountered relationship characterising mappings from AAT, LCSH and MeSH, and the second most frequently occurring in the case of UNESCO, was due to the use of a universal classification scheme as the target terminology. DDC attempts to provide an epistemological interpretation of knowledge, and the treatment of concepts is therefore often more broad, even when analytico-synthetic features are employed. The mapped schemes’ terms tend to be more granular as indicated by the proportion of species-genus relationships. The infrequent assignation of reciprocal match code 6

(genus-species or broader term match) in our study appears to support this assertion. Code 5 was the second most frequently assigned in the case of UNESCO, although the first and second place rankings only differed by a single mapping. The inability of DDC to match the granularity of the mapped schemes is telling and suggests that in many cases the target will actually degrade the signal for the user [15, 18], as discussed in section 2. Furthermore, the nature of DDC as a bibliographic classification scheme dictates that it is complex when compared to relational vocabularies or term lists, and is often not conducive to term-to-term mappings. Analytico-synthetic features have to be regularly employed to express particular concepts and therefore concepts do not exist in a formal sense. This renders the identification of direct equivalence problematic. It is therefore important to note that although the use of target schemes (e.g. DDC) often proves advantageous [3, 21, 22] and theoretically sound [19, 20, 46], their use may actually compromise retrieval performance for users.

Concept match (code 14) proved the second most frequently assigned match type in characterising relationships between terms mapped from AAT and LCSH, and joint second (with exact match) for MeSH. For UNESCO code 14 was the most frequently applied. This indicates that concept matches are evident across all schemes and are a necessary means of identifying like terms. This assertion holds when considering both general and discipline-specific schemes, indicating a good degree of conceptual equivalence across all schemes and accounting for 19.5% of total equivalence. Closer examination reveals a higher proportion of conceptual equivalence between universal schemes and DDC (i.e. LCSH (24%); UNESCO (30%)) than evident in the case of more granular terminologies such as AAT (14%) and MeSH (10%).

Code 9 (further specification) ranked the third most frequently assigned match type when characterising relationships between terms from LCSH and DDC. We consider this to be a consequence of the structural nature of LCSH. For example, LCSH is a relational vocabulary (i.e. subject heading list) employing the use of subdivisions. Where the DDC hierarchy reads 'Computer programming, programs, data > Programming > Programming languages' the equivalent LCSH heading would read 'Programming languages (Electronic computers)'. While the DDC hierarchy provides contextual information clarifying that the programming languages being referred to directly relate to computers, the lack of sufficient hierarchical semantic structure in LCSH necessitates the use of qualifiers (i.e. 'Electronic computers'), thus providing 'further specification'.

Aside from the three most frequently assigned match types (narrower term, concept match and exact match), which the data highlight as characteristic of frequently occurring relationships between terms in disparate schemes, and with the exception of code 6 (broader) as a reciprocal entity of narrower, it appears that the remaining codes assigned (3, 4, 7, 9 and 10) (Table 6) can each be considered as a form of exact or concept match. This would suggest that the 9.5% of match types defined by codes 3 (exact match with intervening

characters), 4 (plural form), 7 (part of speech difference), 9 (further specification) and 10 (spelling variation) could be combined and considered more generally as exact or concept matches. It is considered unlikely that the user of a terminology server would benefit from the knowledge that, for example, ‘absenteeism (labor)’ has further specification than ‘absenteeism’. Users simply want to know that the terms show some level of equivalence in respect of the concepts they represent [47] unless, of course, the further specification alters the context of the term.

7.3. Match types validated

Recall that the primary aim of this study was to test the hypothesis that such a large number of match types – across a variety of terminologies and using Chaplan's rules of application – is unnecessary and could be collapsed into a smaller (perhaps more manageable) number. The present study validated the application of nine of Chaplan’s 19 match types as detailed in Table 10.

Match Type Code	Mapping Type
1	Exact Match
3	Exact Match with Intervening Characters
4	Plural Form
5	Species-Genus Subordination
6	Genus-Species Superordination
7	Part-of-Speech Difference
9	Further Specification
10	Spelling Variation
14	Concept Match

Table 10. Match types verified.

All four mapped schemes demonstrated incidences of exact matches, species-genus subordination and concept matches. This suggests that these three forms of equivalence should be retained in any future set of mapping types proposed. Exact match, narrower (and broader) terms and concept match all constitute benefits for the retrieval of information since they provide the user with further information on a subject area and potentially relevant terms with which to search. In addition, code 3 (exact match with intervening characters) was assigned

to mapped terms originating from MeSH and UNESCO; code 4 (plural form) was applied to one term mapped from LCSH to DDC; code 6 (genus-species subordination) was verified by terms from LCSH and UNESCO; code 7 (part-of-speech difference) was assigned to a mapping from UNESCO; code 9 (further specification) applied to relationships between terms from LCSH, MeSH and UNESCO; code 10 (spelling variation) proved valid in characterising the association between one UNESCO term and a DDC equivalent. A closer examination of these match types suggests that they are not sufficiently distinct to warrant their inclusion in a reduced set of mapping types, with the possible exception of code 6. Code 6 is likely to be more frequently assigned should a scheme with extremely broad subject groupings and a low level of granularity be mapped to DDC. In other words, code 6 could prove valid when a scheme contains top terms exemplifying a broader subject scope than those contained within the target terminology.

Combining match types, where frequency counts are low and/or scheme-specific, is proposed thus reducing the overall range of mapping types required within a terminology service. Code 3 is essentially an exact match and it is proposed that such cases be characterised accordingly. The addition of e.g. a space, a hyphen or a colon does not sufficiently change the meaning of a term to warrant the need for an additional match type. It is proposed that codes 4, 7, 9 and 10 constitute a form of concept match and, as such, should be assigned code 14. In each of these cases mapped terms convey equivalent concepts.

7.4. Extraneous match types

Based on the results shown in Table 6, codes 2, 8, 11, 12, 13, 15, 16, 17, 18 and 19 appear redundant in the case of the current schemes and term sets extracted. This is not safe to assume at this stage however since the methodology led authors to actively seek exact matches, or as near to an exact match as possible. As a result, it is probable that additional match codes are required depending on user circumstances. For example, it is possible that in a search for ‘employment’, ‘unemployment’ may be a more useful term than ‘work’. It follows that more of Chaplan’s match types than indicated by the data above may be relevant and that the set of match types presented in section 6.4 should be supplemented accordingly. Further research is required in this area to determine whether or not selected match types that appear extraneous in the present study, but were proven necessary within Chaplan’s study, may in fact prove valid to the user in specific scenarios. However, the current authors would argue that codes 2, 8, 11, 12, 13, 16 and 17 constitute forms of concept match. In the case of code 15 (homograph), it is questionable that any such form of relationship should be imposed between terms. Although homographs appear as exact matches on a presentational basis, their meanings are not equivalent. It follows that no level of exact or partial match is relevant and that such terms essentially constitute a ‘no match’ relationship. In the current study, conducted within the context of an M2M terminology server employing a DDC spine, such terms would be presented to the user within their DDC hierarchies, enabling their sense of meaning to

become apparent. The process of user disambiguation should be sufficient to handle any potential confusion over the sense of homographs. Chaplan herself did not find any incidence of match code 17 (date or numerical variation), making justification for the inclusion of this type of equivalence in her set of 19 unclear. This leaves match code 18 (no match), the relevance of which within an intellectual mapping scenario is questionable.

8. Conclusion and further work

The present study has confirmed that Chaplan's set of 19 match types is excessive for the purpose of characterising equivalence relationships between terms in disparate schemes within the context of a DDC-spine based terminologies server, when examined in relation to a subset of terms extracted from AAT, LCSH, MeSH and UNESCO. A total of nine of the 19 equivalence relationships were verified, with exact, concept and species-genus subordination proving the most frequently encountered types across all four schemes. This supports our stated hypothesis and provides us with a generic suite of match types.

It is considered likely that the nine match types verified from Chaplan's set could be further reduced, provided they are sufficiently well defined, to form a set closer to that proposed by the conceptually based SKOS Core MVS model. The present study indicates that the developing SKOS Core MVS is insufficient as it stands and requires modification, since currently only three match types (exactMatch, broaderMatch and minorMatch) appear applicable in the context of a distributed terminology server. This is consistent with the work of Liang *et al.* [2005] who found that the MVS required supplementing and redefining to sufficiently express match types identified between AGROVOC and the Chinese Agricultural Thesaurus. Nevertheless, the value of either approach for users engaging in a process of disambiguation remains unclear. We propose to conduct an appropriate user study to verify that a conceptual basis for match types is sufficient for the purposes of retrieval via a terminology server and that lexical differences do not compromise user success in this context.

Possible limitations of the present study have been noted. The choice of schemes investigated together with the extraction of terms to be mapped (although random) may have affected the outcome. In order to eliminate any such bias, the study should be extended to look at a wider range of schemes and a greater selection of terms from each. In addition, encoding of match types was conducted by two individuals; the effect of a larger number of encoders would be valuable to observe. Such limitations will be considered in future work, as described above.

9. References

- [1] Lois Mai Chan and Marcia Lei Zeng, Ensuring interoperability among subject vocabularies and Knowledge Organization Schemes: a methodological analysis, *IFLA Journal* 28(5/6) (2002) 323-327.
 - [2] Marcia Lei Zeng and Lois Mai Chan, Trends and issues in establishing interoperability among Knowledge Organization Systems, *Journal of the American Society for Information Sciences and Technology* 55(5) (2004) 377-395.
 - [3] Traugott Koch, Heike Neuroth and Michael Day, Renardus: Cross-browsing European subject gateway via a common classification system (DDC). In: I.C. McIlwaine (ed), Proceedings of the IFLA satellite meeting held in Dublin, Ohio, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC (K.G. Saur, Munchen, 2003) 25-33.
 - [4] Martin Doerr, Semantic problems of thesaurus mapping, *Journal of Digital Information* 1(8) (2001). Available at: <http://jodi.tamu.edu/Articles/v01/i08/Doerr/> (accessed 10 July 2006).
 - [5] Ceri Binding and Douglas Tudhope, KOS at your Service: Programmatic Access to Knowledge Organisation Systems 4(4) (2004). Available: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Binding/> (accessed 10 July 2006)
 - [6] Jens-Erik Mai, The future of general classification, *Cataloging & Classification Quarterly* 37(1/2) (2003) 3-12.
 - [7] Diane Vizine-Goetz, Carol Hickey, Andrew Houghton and Roger Thompson, Vocabulary mapping for terminology services, *Journal of Digital Information* 4(4) (2004). Available at: <http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/> (accessed 10 July 2006).
 - [8] Margaret A. Chaplan, Mapping Laborline Thesaurus terms to Library of Congress Subject Headings: implications for vocabulary switching, *Library Quarterly* 56(1) (1995) 39-61.
 - [9] H. H. Neville, Feasibility study of a scheme for reconciling thesauri covering a common subject, *Journal of Documentation* 26(4) (1970) 313-336.
 - [10] June P. Silvester and Paul H. Klingbiel, An operational system for subject switching between controlled vocabularies, *Information Processing & Management* 29(1) (1993) 47-59.
 - [11] E. Freyre and M. Naudi, MACS: Subject access across languages and networks. In: I.C. McIlwaine (ed), *Proceedings of the IFLA satellite meeting held in Dublin, Ohio, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC* (K.G. Saur, Munchen, 2003) 3-10.
 - [12] Genevieve Clavel-Merrin, MACS (Multilingual access to subjects) A virtual authority file across languages, *Cataloging & Classification Quarterly* 39(1/2) (2004) 323-330.
-

- [13] Emma McCulloch, Ali Shiri and Dennis Nicholson, Challenges and issues in terminology mapping: a digital library perspective, *The Electronic Library* 23(6) (2005) 671-677.
- [14] Verina Horsnell, Intermediate lexicon for information science: a feasibility study (Polytechnic of North London, London, 1974).
- [15] Verina Horsnell, The Intermediate Lexicon: an aid to international co-operation, *Aslib Proceedings* 27(2) (1975) 57-66.
- [16] R. T. Niehoff, Development of an integrated energy vocabulary and the possibilities for online subject switching, *Journal of the American Society for Information Science* 27(1) (1976) 3-17.
- [17] Patricia S. Kuhr, Putting the world back together: Mapping multiple vocabularies into a single thesaurus. In: I.C. McIlwaine (ed), *Proceedings of the IFLA satellite meeting held in Dublin, Ohio, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC* (K.G. Saur, Munchen, 2003) 33-42.
- [18] E. J. Coates, Switching languages for indexing, *Journal of Documentation* 26(2) (1970) 102-110.
- [19] Geoffrey A. Lloyd, The Universal Decimal Classification as an international switching language. In: Hans Wellisch and Thomas D. Wilson (eds.), *Subject retrieval in the seventies: Proceedings of an international symposium held at University of Maryland, Maryland, 14-15 May 1971*, (Greenwood Publishing Company, Connecticut, 1972).
- [20] Elaine Svenonius, Use of classification in online retrieval, *Library Resources & Technical Services* 27(1) (1983) 76-80.
- [21] Dennis Nicholson, Subject-based interoperability: issues from the High Level Thesaurus (HILT) project. In: *Proceedings of the 68th IFLA General Council and Conference – Classification and indexing, Glasgow, UK, August 18-24 2002*, (IFLA, The Hague, 2002). Available at: <http://www.ifla.org/IV/ifla68/papers/006-122e.pdf> (accessed 10 July 2006)
- [22] Maria Balikova, Multilingual Subject Access to catalogues of National Libraries (MSAC): Czech Republic's collaborations with Slovakia, Slovenia, Croatia, Macedonia, Lithuania and Latvia. In: *Proceedings of the World Library and Information Congress: 71st IFLA General Conference and Council – Classification and indexing with cataloguing, Oslo, Norway, August 14-18 2005*, (IFLA, The Hague, 2005). Available at: <http://www.ifla.org/IV/ifla71/papers/044e-Balikova.pdf> (accessed 10 July 2006)
- [23] Ingetraut Dahlberg, Towards establishment of compatibility between indexing languages, *International Classification* 8(2) (1981) 86-91.
- [24] Dennis Nicholson, Alan Dawson and Ali Shiri, HILT: A pilot terminology mapping service with a DDC spine, *Cataloging & Classification Quarterly* 42(3/4) (2006) 187-200.

- [25] Ali Shiri, Dennis Nicholson and Emma McCulloch, User evaluation of a pilot terminologies server for a distributed multi-scheme environment, *Online Information Review* 28(4) (2004) 273-283.
- [26] Dennis Nicholson, HILT: High-Level Thesaurus Project M2M Feasibility Study: final report to JISC, (CDLR, Glasgow, 2005). Available:
<http://hilt.cdlr.strath.ac.uk/hiltm2mfs/0HILTM2MFinalReportRepV3.1.pdf>
- [27] Alistair Miles and Dan Brickley (eds), SKOS Core Guide: W3C Working Draft 2 November, World Wide Web Consortium (W3C) (2005). Available at: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/> (accessed 10 July 2006)
- [28] Alistair Miles and Dan Brickley (eds), SKOS Mapping Vocabulary Specification, World Wide Web Consortium (W3C) (2004). Available at: <http://www.w3.org/2004/02/skos/mapping/spec/> (accessed 10 July 2006)
- [29] SWAD-Europe, Inter-Thesaurus Mapping: A guide to the SKOS-Mapping RDF schema for inter-thesaurus mapping, World Wide Web Consortium (W3C) (2003). Available:
<http://www.w3.org/2001/sw/Europe/reports/thes/8.4/> (accessed 10 July 2006)
- [30] SWAD-Europe. Available at: <http://www.w3.org/2001/sw/Europe/> (accessed 10 July 2006)
- [31] A. Liang, M. Sini, Chang Chun, Li Sijing, Lu Wenlin, He Chunpei and J. Keizer, The mapping schema from Chinese Agricultural Thesaurus to AGROVOC, *6th Agricultural Ontology Service (AOS) Workshop on Ontologies: the more practical issues and experiences, July 25-28, Vila Real, Portugal, 2005*, (Food and Agriculture Organization, Rome, 2005). Available: <ftp://ftp.fao.org/docrep/fao/008/af241e/af241e00.pdf> (accessed 10 July)
- [32] Lois Mai Chan and Diane Vizine-Goetz, Feasibility of a computer-generated subject validation file based on frequency of occurrence of assigned LC Subject Headings. Phase II, nature and patterns of invalid headings, *Annual Review of OCLC Research* 1995 (1996). Available at:
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?objid=0000003370> (accessed 10 July 2006).
- [33] Diane Vizine-Goetz, Popular LCSH with Dewey numbers, *OCLC Newsletter* (233) (1998). Available at:
<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?objid=0000003449> (accessed 10 July 2006).
- [34] Library of Congress, MARC21 concise format for authority data (2005). Available at:
<http://www.loc.gov/marc/authority/> (accessed 10 July 2006).
- [35] Joan S. Mitchell (ed), *People, places and things. A list of popular Library of Congress Subject Headings with Dewey numbers* (Forest Press, Ohio, 2001).
- [36] OCLC, WebDewey (2006). Available at: <http://www.oclc.org/dewey/versions/webdewey/> (accessed 10 July 2006).
-

- [37] Miltiadis Lytras, Miguel-Angel Sicilia, John Davies and Vipul Kashyap, Digital libraries in the knowledge era: knowledge management and Semantic Web technologies, *Library Management* 26(4/5) (2005) 170-175.
- [38] Linda Cantara, Encoding controlled vocabularies for the Semantic Web using SKOS Core, *OCLC Systems & Services* 22(2) (2006) 111-114.
- [39] A. C. Liang and M. Sini, Mapping AGROVOC and the Chinese Agricultural Thesaurus: definitions, tools, procedures, *New Review of Hypermedia and Multimedia* 12(1) (2006) 51-62.
- [40] OCLC, Dewey services: Dewey Decimal Classification (2006). Available at: <http://www.oclc.org/dewey/> (accessed 10 July 2006).
- [41] Library of Congress, Library of Congress Authorities (2006). Available at: <http://authorities.loc.gov/> (accessed 10 July 2006).
- [42] United States National Library of Health, Medical Subject Headings (2006). Available at: <http://www.nlm.nih.gov/mesh/> (accessed 10 July 2006).
- [43] UNESCO and the University of London Computing Centre, UNESCO Thesaurus (2002). Available at: <http://www2.ulcc.ac.uk/unesco/> (accessed 10 July 2006).
- [44] J. Paul Getty Trust, Art and Architecture Thesaurus Online (2000). Available at: http://www.getty.edu/research/conducting_research/vocabularies/aat/ (accessed 10 July 2006).
- [45] OCLC, Introduction to Dewey Decimal Classification (OCLC, Ohio, 2006). Available at: <http://www.oclc.org/dewey/versions/ddc22print/intro.pdf> (accessed 10 July 2006).
- [46] Songqiao Liu, Decomposing DDC synthesized numbers, *Proceedings of the 62nd IFLA General Conference, August 25-31, Beijing, 1996*, (IFLA, The Hague, 1996). Available at: <http://www.ifla.org/IV/ifla62/62-sonl.htm> (accessed 10 July 2006)
- [47] Robert Krovetz and W. Bruce Croft, Lexical ambiguity and information retrieval, *ACM Transactions on Information Systems* 10(2) (1992) 115-141.