# UDC in Action

Richard Smiraglia
*School of Information Studies, University of Wisconsin, Milwaukee, USA*

Andrea Scharnhorst
*Data Archiving and Networked Services, Royal Netherlands Academy of Art and Sciences, The Hague & Amsterdam, The Netherlands*

Almila Akdag Salah
*University of Amsterdam, The Netherlands*

Cheng Gao
*Dalian, China*

**Abstract:** The Universal Decimal Classification (UDC) is not only a classification language with a long history; it also presents a complex cognitive system worthy of the attention of complexity theory. The elements of the UDC) scheme: main numbers, auxiliary numbers and connecting symbols are combined into symbolic strings, which in essence represent a complex networks of concepts. This network forms a backbone of ordering of knowledge and at the same time allows expression of different perspectives on various products of human knowledge. In this paper we look at UDC strings taken from library catalogues. Specifically, we analyse classmarks from the UDC authority file of the University Library in Leuven and an extraction of UDC numbers from the OCLC WorldCat. Comparing those sets with the UDC Master Reference File (UDC MRF), we look into the length of strings, the occurrence of different auxiliary signs, and the resulting connections between UDC classes. We apply methods and representations from complexity theory. Mapping out basic statistics on UDC classes as used in libraries from a point of view of complexity theory bears different benefits. Deploying its structure could serve as an overview and basic information for users among the nature and focus of specific collections. A closer view into combined UDC numbers reveals the complex nature of the UDC as an example for a knowledge ordering system, which deserves future exploration from a complexity theoretical perspective.

**Keywords:** bibliographical data; metadata; classification; UDC; Universal Decimal Classification; parsing; statistical analysis; complexity theory; visualization

## 1. Introduction

Inherent in all knowledge domains is a certain systematization of knowledge about phenomena. But, as Kedrov (1975) expresses it, the need to classify all of the knowledge of humankind regularly emerges. This is the domain in which universal classifications, such as the Universal Decimal Classification (UDC), operate (Slavic, 2008; 2008b). The design of classification systems used to order knowledge and retrieve information entails a wealth of contextual information about concepts: which are thought to be most relevant, what is the envisioned relationship between them, and how best to express this in a formal way. Such system designs can be approached as though they were complex phenomena of human knowledge production. In this paper we apply some of the typical methods for studying complex systems based on quantitative information that can be gathered about them.

Classification research is the scientific field where traditionally form, structure, meaning, and increasingly also the evolution, of classification systems are studied and implications for information retrieval are discussed (Tennis, 2006; 2007). Still in this field, quantitative methods - not to mention large-scale data analysis and accompanying visual explorations (Osinska, 2010; Börner, 2010) are not yet standard elements in the academic discourse. Digital representations of classification systems, UDC Master Reference File (UDC MRF) being one such example, and their instantiations in collections of museums and libraries (and their online catalogues) allow and call for computer-based methods of analysis. Moreover, quantitative indicators about the size, inner composition, growth and change of a classification system are important for the monitoring of the design of the system as well as of its application. So, it is no surprise that this kind of information is collected and analysed, for instance, by the UDC editors in the process of the maintenance and revision of UDC (McIlwaine, 2007). However, the main focus of makers and users of the UDC, as of the makers of other classification systems, is not the scientific analysis of this system on a fundamental level. This might explain the lack of systematic quantitative studies about these systems. For outsiders, on the other hand, it is not always easy to have access to numeric information about the implementation of the UDC, and comparable systems.

In earlier studies the authors of this paper have used the annual file releases of the UDC MRF, a printed early UDC editions (from which data were entered manually) and information from secondary sources[1]. These were used to display, analyse and discuss structure and evolution of the UDC classes and the use of common auxiliary numbers in the UDC (Scharnhorst et al., 2011; Akdag Salah et al. 2012). In an earlier comparative study of UDC and Wikipedia category structures (Akdag Salah et al., 2011; 2012a) it was evident that to some extent apples and oranges were compared: the Wikipedia is a collection containing an emerging category system, while the UDC MRF represents a stand-alone categorization i.e. terminological tool in its own right which is not directly connected to information resources it is designed to index. This was the reason for the authors to look into the actual UDC application data.

This paper presents results from the study of two sets of UDC numbers retrieved from two bibliographic systems containing UDC data. Courtesy of colleagues from Katholieke Universiteit Leuven (K. U. Leuven) and OCLC, we received a sample of UDC numbers from the K. U. Leuven library catalogue and WorldCat - both from May 2011. In this paper, we demonstrate how insights based on statistical analysis and corresponding visuals can help us understand not only the envisaged but also the actual use of a classification system.

1 For instance, *The Universal Decimal Classification - a guide to its use* (McIlwaine, 2007) and annual issues of *Extensions & Corrections to the UDC*.

## 2. Datasets

### *2.1. OCLC dataset*

The OCLC dataset is an ASCII text file consisting of 9,055,623 entries extracted from 214,596,487 bibliographic records using the "080" field in WorldCat. After the non-UDC number records (either those that are missing data in the UDC number field or where the UDC number field was missing): were taken out, there were 8,944,669 records in total. Of these records another 570,629 have been discarded as non-UDC numbers. Eventually, we identified 8,374,040 entries containing UDC numbers both simple and pre-combined. As shown in Table 1, each line in the data file contains two columns. The first column contains record ID and the second column holds UDC numbers, preceded by a subfield tag). Each row i.e. record entry represents a document with an assigned UDC number

**Table 1:** *Sample from the original file of extracted numbers from OCLC WorldCat*

| | |
|---|---|
| 16213477 | a621.315.2:678.742:621.395 |
| 17323477 | a614.25 |
| 17343477 | a618.177:616.697 |
| 36603477 | a82-31/-32 |
| 36613477 | a597 |
| 36613477 | a82-93 |
| 36663477 | a614.253.5.001 |
| 37253477 | a82-93 |
| 37253477 | a820(73)-31 |
| 37963477 | a577 |

### *2.2. K. U. Leuven library dataset*

The set from KU Leuven library is an export of the UDC authority file from their library system (DOBIS/LIBIS). The file contains UDC classmarks used in the university library of Leuven. The original file has 95,544 lines. In Table 2 below, we can see a section of the export: the first column contains a string with the structure $$8 UDC number $$a UDC number's descriptor (i.e. number described by words) and $$9 at the end indicates the language of the caption. The second column indicates how often this UDC number is used in bibliographic records in the library system. In the example shown in Table 2 we can see on the first line a UDC number "18" which is a common auxiliary of time denoting 19th century, followed by the caption "19e eeuw. Periode 1800-1899" describing the UDC code in Dutch. This auxiliary can be found twice in the bibliographic records in the library system.

**Table2:** *An excerpt from the KU Leuven library datase, first 10 lines*

| | |
|---|---|
| $$8"18"$$a19e eeuw. Periode 1800-1899$$9dut | 2 |
| $$8(043)U1$$aDissertaties--U1$$9dut | 6 |
| $$8(043)U2$$aDissertaties--U2$$9dut | 3 |
| $$8(043)W1$$aDissertaties--W1$$9dut | 6 |
| $$8001 <03>$$aWetenschap en kennis--(algemeen)--Naslagwerken. Referentiewerken$$9dut | 3 |
| $$8001 <061>$$aWetenschap en kennis--(algemeen)--?<061>$$9dut | 1 |

Specific to the K.U. Leuven library is the use of authority file with UDC codes to derive and control natural language terms and support multilingual searching and subject browsing without users being aware of classification. UDC authority file in the K. U. Leuven library system is created to support searching and browsing of UDC numbers either by using French or Dutch search terms or by numbers (c.f. Schallier, 2004; 2004a)

> "In conjunction with LIBIS, our IT-department we developed an interface that can handle a search by UDC words as well as by codes, all taken from our UDC authority file. Both words and codes can be truncated and can be combined with Boolean operators. Thus when we enter the search statement "Leuven?", we find 100 relevant UDC codes ... . It is possible to refine the results by selecting one of the 10 main UDC classes ... or by combining with another search term... efforts have been made to translate the Dutch descriptors (preceded by n-) into French (preceded by f-). Classes 5 en 6 (exact and applied sciences) are entirely in English. Although the presentation still needs work, it already reveals the multilingual potential of our authority file. " (Schallier, 2004a: 22)

A bibliographic record from their OPAC appears as follows:

| | |
|---|---|
| Title | Social structure and change : Finland and Poland : comparative perspective / Ed. by Erik Allardt and Wlodzimierz Wesolowski. |
| Publ. Year | 1978 |
| Publisher | Warszawa : Polish scientific publ., |
| Editor(s) | Allardt, Erik / Wesołowski, Włodzimierz / |
| Physical details | 391 p. |
| Subjects | Social change. Sociale ontwikkeling. Sociale veranderingen. Modernisering. Evolutie .Sociale revolutie. Modernisme : 316.42<br>Sociale structuur --(sociologie) --Finland : 316.3 <480><br>Sociale structuur --(sociologie) -- Polen : 316.3 <438> |

This bibliographic description would be linked to two UDC authority file entries. Rather than representing the actual documents the file we work with represents the authority list of UDC codes and their respective natural language descriptors with an information of the frequency of their use within the library catalogue. We ignored the frequency data (second column in Table 2), and analysed and aggregated the unique occurrences. Eventually, for K. U. Leuven 91,132 lines with UDC numbers were analysed.

In addition to two bibliographical data sets we also used the UDC MRF 2008.

### 3. Data processing

Once the UDC numbers are extracted from the datasets, they need to be parsed in order to analyse them further. UDC numbers are strings, which contain numeric characters as well as other specific symbols. In the parsing attention must be paid to the sequence of numeric and non-numeric symbols. For example, the OCLC dataset contains the string:

**394.4 :[92(100+437):329(437).15(091)+327.32(100)]**

First, **394.4** is a main UDC number standing for "Public ceremonial, coronations," and colon **:** is a connecting symbol representing "simple relation". Square brackets are used for subgrouping. Everything within the **[....]** brackets is a unity. This unity starts with another main UDC class number **92**, standing for "Biographical studies. Genealogy. Heraldry.Flags". The () parentheses when starting with a non-zero numeric character denote a common auxiliary number of place. In our case **(100+437)** indicates "(100) All countries in general" AND "(437) Czechoslovakia (1918-1992)". The next part of this complex combination in the square brackets starts again with UDC number **329**.15 standing for "Political parties with a communist attitude" the place auxiliary of place **(437)** Czechoslovakia (1918-1992)" is intercalated between **329** and **15** to allow for collocation of all Czechoslovakian parties irrespective of their political orientation, and then ordered by a type - thus entire number represents a topic "Communist party of Czechoslovakia" which is then further specified by a common auxiliary of form (091) denoting presentation in a historical form to express "the history of communist party of Czechoslovakia". The last part in the unity starts with a "+" sign, the common auxiliary sign for addition/coordination introducing the next UDC number combination in the string consisting of two parts: "327.32 International solidarity of the working class" and "(100) All countries in general." For an expert the UDC string above clearly points to a work dealing with:

*Public celebrations/ceremonies with significant biographical and historical elements, or even artifacts to do with celebrations (e.g. flags, banners) and which involve historical personalities (both Czechoslovakian and international) linked to the history of Czechoslovakian Communists Party and international movement of solidarity of the working class - in the world. Such a book would probably have something to do with parades and celebrations of May 1 International Workers' Day or similar events in former Czechoslovakia.*

For the intended statistical analysis, we calculated the length of the strings, but also extracted information about which UDC class occurred in combination with which other class. We also differentiated between the different operations possible for combining UDC numbers. So, for the later analysis, the number above would deliver a tick in a matrix of row and columns standing for the UDC classes, between class 3 and class 3,9,3 taking account of the ":" operation of relation. It would also contribute a tick between class 3 and class 3 because of the "+" operation. But no operation would be counted for the (100+437) part, because in this case it would be an operation among common auxiliaries of place, and not among classes. This illustration also shows that operation based on parsing can easily go wrong if any symbols are ignored. At the beginning of our experiments we received quite a number of entries for operations that seemed to come from class 4. Knowing that class 4 is empty and not used in the current UDC, we went back to the parsing and found that all 4xx cases actually emerged from auxiliaries of place.

UDC numbers when applied in collection indexing may contain local extensions and symbols other than those present in the scheme. For instance, <> brackets sometimes are used to add other notation or information to the UDC number and can also be used to isolate elements of UDC numbers for the purpose of searching. We ignored information that appeared in those brackets. The sign of asterisk * is used in UDC to introduce non-UDC notations, and therefore it was equally ignored in our parsing.

## UDC in Action I - profile of collections

Based on the OCLC set and the Leuven set, we can now compare the actual use of the UDC in bibliographic data and in the UDC authority file with the content of the UDC MRF file (2008 version). There is one more problem to address before we can turn to analysing the distributions themselves. Both the OCLC and the Leuven set contain strings of simple and combined UDC numbers which may consist of:

- combination of two or more simple main numbers connected with symbols /, + or : e.g. **394.4:92**
- combination of simple main numbers combined with common auxiliary numbers (in this case there are no connecting symbols) e.g. **32(100)**
- combination of two or more common auxiliary number using symbols which represent common auxiliary signs in conjunction with UDC numbers e.g. (437+100)
- combination of main numbers with special auxiliary numbers
- combination of common auxiliary numbers with special auxiliary numbers

The first case can be understood as operations (similar to algebraic operations). For the sake of simplicity, we decided to only analyze the links (combinations) between UDC classes. We ignored common auxiliaries (general concepts of language, place, time, form, materials etc.) and special auxiliaries (specific subdivision within main tables) and combinations involving special and common auxiliariesIn all other cases, a combination with a common or special auxiliary occurs. Due to limitations in this research, these types of combinations are not included in the analysis. Hence, in the network analysis common auxiliaries (general concepts of language, place, time, form, materials etc.) and special auxiliaries (specific subdivision within main tables) were not taken into account.

For the distribution of UDC numbers across the 9 main classes we applied two methods of counting. For the first method, we only used the UDC number at the beginning of the string. In the second method of counting, each UDC number (before and after an auxiliary sign) was counted. So, if we look at the arbitrary string: 123.456:213.465. From this string the first counting method would deliver one occurrence of class 1. With the second counting method both class 1 and class 2 would receive a tick of occurrence. Concerning the class distribution there is no big difference between the methods, so here we only display results based on counting method 1.

The MRF set of 2008 is with 68,195 i.e. 54,980 when common auxiliary tables where taken out records the smallest in absolute terms, followed by Leuven with 91,132 records. The OCLC set with 8,374,040 analysed lines is three orders of magnitude larger.

We have argued earlier that the UDC MRF to a certain extent reflects how changes in the sciences are perceived and accommodated in the editorial processes. The primary use of the UDC for academic libraries explains the relatively large size of class 6. In comparison to the UDC authority file from K. U. Leuven library system which is dominated by Social Sciences (3), followed by Religion (2), and Applied Sciences (6) in third place. In the OCLC dataset, Social Sciences (3) are leading, followed by Applied Sciences (6) and Language, Linguistics and Literature (8).

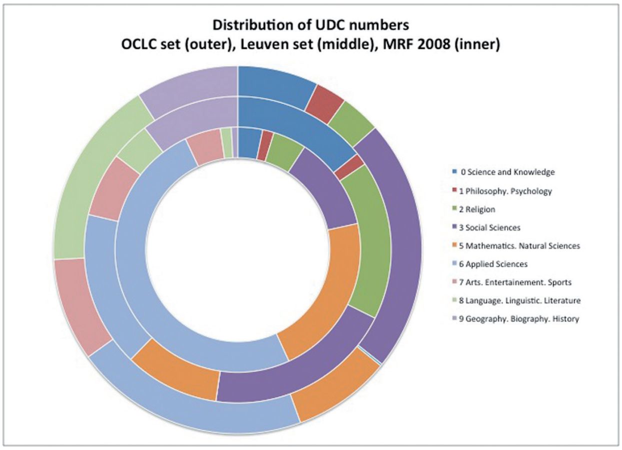



**Figure 1:** *Distribution of UDC numbers across classes, in design as well as in practical use*

The actual use of UDC numbers mirrors the content of library collections (or an agglomeration of them as is found in OCLC). Not surpringsingly actual use deviates from the MRF, illustrating the collective diverse foci of collections.

## UDC in Action II - analysis of the complexity of the UDC

### *Distribution of UDC strings*

In the previous section we already pointed to the complexity of the strings appearing in bibliographic records using UDC numbers, and the challenges for parsing. The UDC is a complex language that can capture complex relationships between concepts rather than a static reference system in which concepts can be easily and uniquely placed. From the point of view of complexity, we took the length of a string as an indication of the complexity of a UDC number expressing the extent of recombination of concepts. The MRF file contains not only single UDC numbers, but different combinations of them leading to strings longer than six characters which appears to be the most frequent length of a single UDC number.

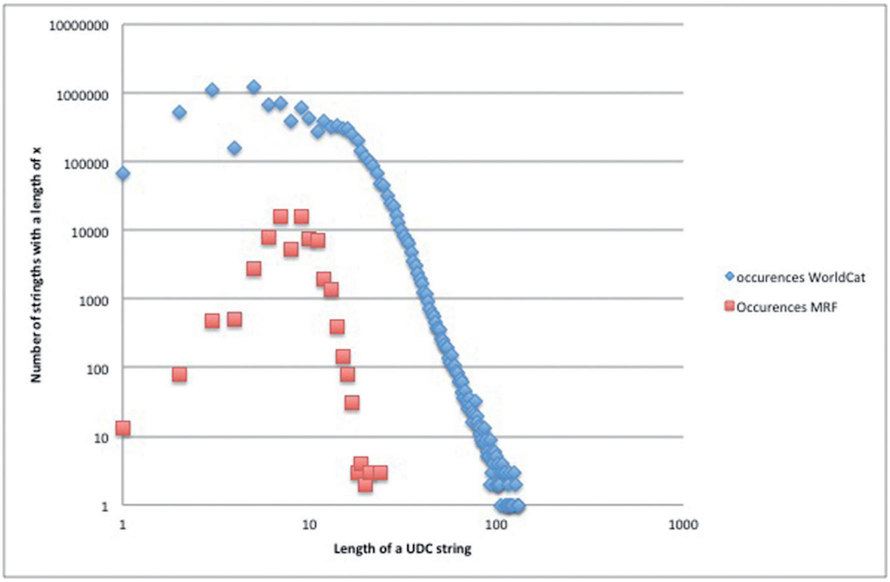



**Figure 2:** *Frequency distribution of UDC strings of a certain length*

How do we read the graphics in Figure 2, which is a typical representation in statistical physics called a distribution function? The first data point (of the bottom curve) represents all strings in the UDC MRF of size 1. There are 68 occurrences of strings of the length 1. Among them are the 9 main classes, but also other signs.

Please do note that both x- and y-axe use a logarithmic scale. The distribution of string length in the MRF file peaks at length 6 (the length of single UDC numbers). But shorter or longer UDC numbers are not normally distributed (in a Gaussian curve) around this expected most frequent length. As in many other complex socio-technical systems (Scharnhorst, 2003) the distribution function is skewed, relatively long strings appear less frequently, the maximum string length in the UDC MRF 2008 is 16. If we look into the same distribution for the OCLC data set, we immediately see that in actual use much more complex UDC numbers occur. They fall in the long tail of the distribution and can extend up to length 135. At the very least this distribution indicates that the UDC as a classification system shows features of a complex and non-linear system.

### 1. Weaving meaning from operations on classes - network views of the UDC

The UDC knows six connecting symbols (or 'relators') the description of which can be found in Tables 1a, 1b and 1h.

- Table 1a contains + *Coordination. Addition (plus sign)* and */ Consecutive extension (oblique stroke sign)*. 1/5=1+2+3+4+5 / = range ….The second operation allows reducing redundancy when combining UDC numbers. For example, the operation 629.734/.735 is equivalent to 629.734+629.735. In general the Table1a auxiliary signs are used to extend subjects.
- Table 1b contains *: Simple relation (colon sign); :: Order-fixing (double colon sign); [] Subgrouping (square brackets)*. Different from Table 1a, Table 1b auxiliary signs are used rather to restrict subjects by defining relations among them.[2]
- Table 1h contains two signs which indicate non UDC notations: *Introduces non-UDC notation (asterisk)*; and *A/Z Direct alphabetical specification*.

In this paper we restrict the analysis to the operations "+" (plus), "/" (stroke) and ":" (colon).

To analyse which kind of operations, or in other words to discover which auxiliaries and auxiliary signs are responsible for the length or complexity of UDC numbers, we parsed the OCLC and the Leuven data set and constructed matrices around common auxiliary signs. As discussed previously in the section on data processing, the parsing is not always straightforward, and in the case of large data sets manual cleaning is out of the question. For control, we kept the class 4 (which is vacant i.e. not used in UDC) in the matrix. Cell values for class 4 indicate the level of noise and are marked in grey.

The sequence in which UDC classes are combined usually has meaning. So, all matrices are non-symmetrical and the related networks are directional. For example, the UDC number *022:11.203:042* is recorded in the matrix_colon table as a link between class 0 and class 1, but also in the same matrix_colon table as a link between class 1 and class 0. The combined number *022:11.203+11.204* contributes one tick to the cell {row class 0, column class 1} in the matrix_colon, and in the matrix_plus between row class 1 and column class 1. Combinations

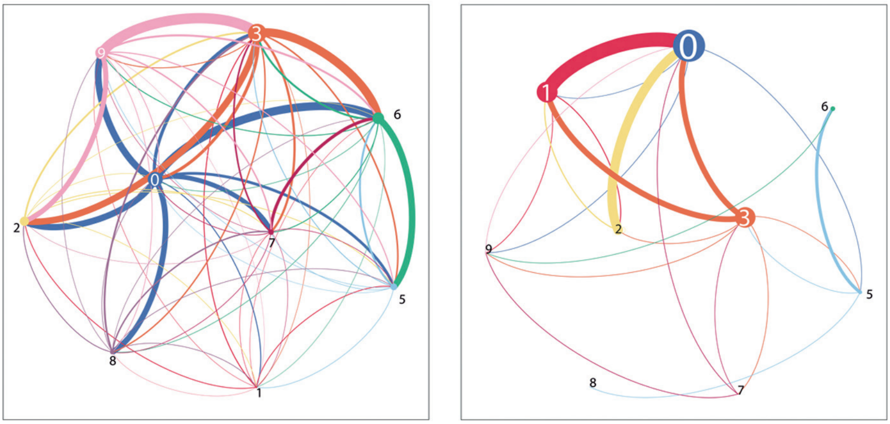
2 See UDC Summary at http://www.udcc.org/udcsummary/php/index.php.

between auxiliaries are not taken into account. Table 3 in the appendix contains the matrices of auxiliary operations from the Leuven data set. For Leuven we only found the common auxiliaries "/" and ":" among classes. Table 4 contains the matrices of operations "+", "/" and ":" for the OCLC dataset.

The network visualization of these matrices is displayed in Figure 3 (Leuven) and Figure 4 (OCLC). For this we abandoned the class 4. The weight of links between classes is indicated by the width of the lines. The size of the nodes corresponds to the relative weighted degree of a node, and indicates the extent to which a class is linked up with other classes. We have not normalized the weight of links by the number of occurrence of a class in the dataset.

In the Leuven dataset, we found no occurrence of the use of the "+" Addition operation between classes. The Consecutive extension "/"can be found inside classes. This is indicated by the self-loops to one node. But the operation is also very popular in links between classes 0, 1, 2 and 3 (Fig 3b). The Simple relation ":" - equally often applied in total, is much more evenly used among all UDC classes (Fig 3a). In the OCLC dataset, we see the operation "+" (Addition) occur mostly in classes such as class 6, 3 and 8 (see Table 4). But, those classes are relatively large. So, one could also apply a normalization to the link weight based on the absolute occurrence of a class in a dataset. The Consecutive extension ("/") occurs one order of magnitude more frequently than "+" Addition. It is also clearly an operation that runs mostly within a class (Fig 4b). But the most popular operation among the three is "Simple relation." This operation (":") is most prominent among class 0 and 8 as can been seen both visually as examine the corresponding matrix (Fig 4a).

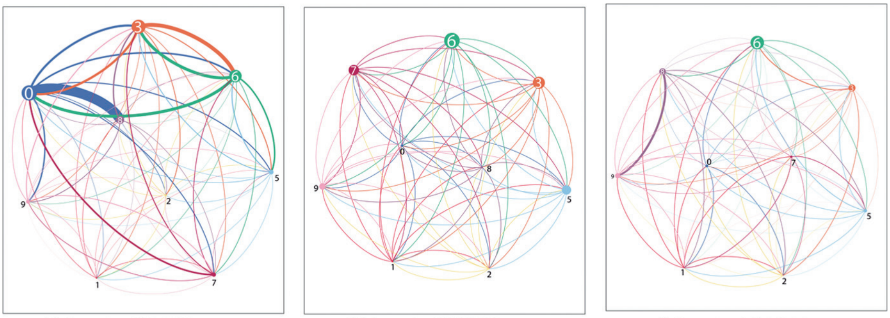
In any case a discussion of the results is humbled by the lack of information we had at hand about the content expected to be large once applied between the classes.




(a) Operation ":", Leuven Data set

(b) Operation "/", Leuven Data set

**Figure 3:** *(a) Network of UDC classes in the Leuven dataset by the use of the operation (a) ":" (Simple relation), (b) "/" (Consecutive extension)*



(a) Operation ":", OCLC Data set (b) Operation "/", OCLC Data set (c) Operation "+", OCLC Data set

**Figure 4:** *Network of UDC classes in the OCLC dataset by the use of the operation (a) ":" (Simple relation), (b) "/" (Consecutive extension), (c) "+" (Addition)*

## Conclusion

In this paper we present different methods of analysing sets of UDC numbers retrieved from library collections. We argue that quantitative methods and related visualizations can be used to compare different instances of the use of the UDC and also to compare them with the designed system. The UDC represents a complex system of symbolic annotation to codify description of the content of documents. A closer view into pre-combined UDC numbers reveals the complex nature of the UDC itself, which deserves future exploration from a complexity theoretical perspective. The main classes form a hierarchical network. Each auxiliary sign creates additional links between classes. The auxiliaries themselves introduce another kind of nodes into this complex network. At the end the UDC presents itself as a network with different kind of nodes and links. In this paper only a selection of those network-representations are displayed. To raise awareness of the UDC as an object of further study for this kind of research is one aim of this paper.

The interpretation of our results is hampered by the fact that the datasets differ in nature, and are therefore only partly comparable, but also that we have only limited context information for the set of UDC numbers under study. The datasets encompass the UDC Master Reference File (the official, authorised version of the UDC), data from a UDC authority file from a library using UDC for subject indexing (K. U. Leuven), and a selection of UDC numbers from bibliographic records belonging to various collections (OCLC). In the last case, we have no information from which collection, related to which works, and even more importantly from which period of time the UDC numbers stem. Hence, we focus on the description of the methods.

It is possible to envisage that analysis of this kind, and especially if designed to answer more specific questions, may be of interest to those responsible for maintaining and developing UDC scheme as well as those working on collection indexing or UDC authority files management. Distribution of classes and length of notation in UDC MRF may reveal pressure points in the notational systems and may help in planning revision and restructuring of classes that seem to be over-crowded. From our consultations with the UDC Editorial Team we have learned that such analysis is of special value to UDC which has inherited a greater level of detail and specificity in the field of applied sciences from the period (1960-1970s) when it was intensively used in technical, industrial and scientific collections (e.g. classes such as 621.397.132.122 Frequency multiplex system). The presence of long notations can indicate two things to editors: a) the subdivision is very specific (possibly even too specific) and specificity may need reducing and b) the schedule contains enumerated complex subjects presented with a simple notation that may need to be restructured in a faceted fashion. Equally, on the side of UDC application, mapping the size (and growth) of UDC data in bibliographic records may give an indication of subject areas in which there is greater number of topics in a collection and where more specific subject indexing may be required. The analysis of relations between subject areas may also indicate the greater need for connectivity and maybe further refinement when it comes to associative relationships between subject areas.

The analysis of UDC numbers in collections represents feedback to editors about the use of classes. However, in this case, the temporal provenance of UDC numbers deserves special attention. Across the editions of the UDC, not only are UDC numbers added and deleted, they also are shifted (and re-labeled) and recombined, as well as receiving changed descriptions.

We are convinced that mapping out basic statistics on UDC classes as used in libraries could be of interest both for the information professionals using the UDC in their daily work as well as for users, who might profit from an overview about the nature and focus of a specific collection.

## Acknowledgment

We would like to thank Ed O'Neill of the OCLC Office of Research who provided us with the OCLC dataset. We would also like to thank Johan Rademakers and Bart Peeters from KU Leuven who provided the Leuven dataset. Aida Slavic gave comments on the paper, and was an indispensable sparring partner for discussion. We would like to thank Krzysztof Suchecki, who being part of the Knowledge Space Lab project, was involved in earlier discussions. Part of this work has been funded by the Network of Excellence for Internet Science, FP7 – 288021, and by the COST Action TD1210 KnowEscape.

## References

Akdag Salah, A. et al. (2012) The evolution of classification systems: ontogeny of the UDC. In: *Categories, contexts and relations in knowledge organization: proceedings of the Twelfth International ISKO Conference, 6-9 August 2012, Mysore, India*. Edited by A. Neelameghan, K.S. Raghavan. Würzburg: Ergon Verlag. pp. 51-57. Also available at: http://arxiv.org/ftp/arxiv/papers/1204/1204.3769.pdf.

Akdag Salah, A. et al. (2012a) Need to categorize: a comparative look at the categories of Universal Decimal Classification system and Wikipedia. *Leonardo*, 45 (1), pp. 84-85. Also available at: http://depot.knaw.nl/13001/1/Leonardo_akdag_2012_NEED_TO_CATEGORIZE-_A.pdf].

Akdag Salah, A. et al. (2011) Design vs. emergence: visualisation of knowledge orders. Map (Courtesy of The Knowledge Space Lab, Royal Netherlands Academy of Arts and Sciences). In: *7th Iteration: Science Maps as Visual Interfaces to Digital Libraries, Places & Spaces: Mapping Science, 2011*. Edited by K. Börner, M. J. Stamper. Available at: http://scimaps.org/maps/map/design_vs_emergence__127/.

Börner, K. (2010) *Atlas of science*. Cambridge, MA: MIT Press.

Kedrov, B. M. (1975) *Klassifizierung der Wissenschaft*. Berlin: Akademie-Verlag Berlin.

McIlwaine, I. C. (2007) *The Universal Decimal Classification: a guide to its use*. Revised ed. The Hague: UDC Consortium.

Osinska, V.; Bala P. (2010) New methods for visualization and improvement of classification scheme. *Knowledge Organization, 37* (3), pp. 157-172.

Schallier, W. (2004) On the razor's edge: between local and overall needs in knowledge organization. In: *Knowledge organization and the global information society: proceedings of the Eight International ISKO Conference, London (UK), 13-16 July 2004*. Edited by I. C. McIlwaine. Würzburg: Ergon. (Advances in Knowledge Organization 9), pp. 269-280. Also available at: http://hdl.handle.net/10150/105347

Schallier, W. (2004a) What a subject search interface can do? Extensions and Corrections to the UDC, 26, pp. 18-28. Also available at: http://hdl.handle.net/10150/105315

Scharnhorst, A. (2003). Complex networks and the web: insights from nonlinear physics. *Journal of Computer-Mediated Communication*, 8 (4). Available at: http://jcmc.indiana.edu/vol8/issu4/scharnhorst.html.

Scharnhorst, A. et al. (2011). The evolution of knowledge, and its representation in classification systems. In: *Classification and Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar, 19-20 September 2011, The Hague, The Netherlands.* Edited by A. Slavic; E. Civallero. Würzburg: Ergon Verlag. pp. 269-282.

Slavic, A. (2008) Use of the Universal Decimal Classification: a world-wide survey. *Journal of Documentation*, 64 (2), pp. 211-228. Also available at: http://hdl.handle.net/10150/105579. [pre-print]

Slavic, A. (2008a) Faceted classification: management and use. *Axiomathes*, 18 (2), pp. 257-271. doi:10.1007/s10516-007-9030-z. Available at: http://springerlink.com/index/10.1007/s10516-007-9030-z.

Tennis, J. T. (2006) Versioning concept schemes for persistent retrieval. *Bulletin of the American Society of Information Science and Technology*, 32 (5), pp. 13-16. Also available at: http://www.asis.org/Bulletin/Jun-06/tennis.html.

Tennis, J. T. (2007) Diachronic and synchronic indexing: modeling conceptual change in indexing languages. In: *Information Sharing in a Fragmented World, Crossing Boundaries. Proceedings of the 35th Annual Meeting of the Canadian Association for Information Science/L'Association Canadienne Des Sciences De L'information, Montreal*. Edited by C. Arsenault and K. Dalkir. Montreal: Canadian Association for Information Science. Also available at: http://arizona.openrepository.com/arizona/bitstream/10150/105762/1/tennis_2007_dlist.pdf

**Richard P. Smiraglia** is Professor, Information Organization Research Group, School of Information Studies, at the University of Wisconsin, Milwaukee. He has defined the meaning of "a work" empirically, and has revealed the ubiquitous phenomenon of instantiation among information objects. Recent work includes empirical analysis of social classification, and epistemological analysis of the role of authorship in bibliographic tradition. His "Idea Collider" research team is working on a unified theory of knowledge. An Honorary Fellow of the Virtual Knowledge Studio, Amsterdam, he is a collaborating member of the Knowledge SpaceLab effort to map the evolution of knowledge in Wikipedia. He holds a PhD (1992) from the University of Chicago. He is editor-in-chief of the journal Knowledge Organization.

**Andrea Scharnhorst** is Head of research at Data Archiving and Networked Services (DANS) - an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW). She is also scientific coordinator of the computational humanities programme at the e-humanities group (http://ehumanities.nl) of the KNAW. She has published about the transfer of concepts and methods at the interface between physics, information sciences, social sciences and humanities. Her work focuses on modelling and simulating the emergence of innovations (new modes of behaviour and learning, forms of communication, technologies or scientific ideas) in social systems.

**Almila Akdag Salah** holds a BSc in industrial design, an MA in art history from Istanbul Technical University and a PhD from the Art History Department of UCLA. Her research interests are in the area of technoscience art and its place in the art historical canon; citation networks and mapping of the threesemi-relateddiscipline[1]s (cognitive science, visual culture and art history) and their interactions. At the Virtual Knowledge Studio of KNAW she was part of the "Knowledge Space Lab" project, which contributed with a map on Wikipedia and UDC (www.scimaps.org/maps/map/design_vs_emergence__127/) to the Places and Spaces exhibit. Currently she conducts at the University of Amsterdam, her research project "DeviantArt: Mapping the Alternative Art World", supported by a VENI award from the Netherlands Organization for Scientific Research.

**Cheng Gao** got her B.S in Computer Science from Beijing Language and Culture University (BLCU) (2005) and M.S. in Computer Science from Beijing University of Posts and Telecommunications (BUPT) (2009). She worked as Scientific Programmer for the Knowledge Space Lab project of the Royal Netherlands Academy of Arts and Sciences (KNAW).