Wikipedia Category Structure in the Light of Dewey Decimal Classification System: an Evaluation

By
Piyali Ghosh

Thesis submitted for the award of the Faculty of Arts of Jadavpur University in partial fulfilment of the requirements for the degree of **Doctor of Philosophy** in Library & Information Science

Department of Library and Information Science
Jadavpur University
Kolkata – 700032
2024

CERTIFICATE

Certified that the thesis entitled "**Wikipedia category structure in the light of Dewey Decimal Classification system: an evaluation**" submitted by me for the award of the Degree of Doctor of Philosophy in Arts at Jadavpur University is based upon my work carried out under the supervision of **Mr. Sunil Kumar Chatterjee,** Professor of the Department of Library and Information Science, Jadavpur University and that neither this thesis nor any part of it has been submitted before for any degree or diploma anywhere/elsewhere.

Countersigned by the
Supervisor: Prof. Sunil Kumar Chatterjee
Dated:

Candidate: Piyali Ghosh
Dated:

DEDICATION

I dedicate this thesis to my beloved parent, my husband and my beloved teachers of the Department of Library and Information Science, Jadavpur University who have helped me in many ways to fulfill my dream.

ACKNOLEDGEMENT

I wish to express my sincere appreciation to my guide Prof. Sunil Kumar Chatterjee for whose guidance I am able to correctly complete my PhD thesis. By working under him, I have gathered that knowledge which will help me a lot in present and in future.

I would like to thank my other teachers. For completing the thesis, the all of them had helped me by giving their advice.

I wish to thank my department library and the university's central library. The librarian and other library staffs had also helped me by advising and providing the right information sources which was needed for this thesis.

So many people had also guided me in different way and encouraged me as I worked on my thesis. I would like to thank them all.

Finally a big thanks to my parents and my husband whose optimism and encouragement had helped to keep me going and finding the right path.

PREFACE

Encyclopedia as a reference work serves a big role in our day to day life and its journey to serve the civilizations was started from very past in history. Now for the net generations the encyclopedias have acquired a new role by transferring its form printed to online. Some of these are free. Wikipedia is the most popular encyclopedia among these. This encyclopedia is created on free and open source Mediawiki software. This encyclopedia's contents are written by general people and before publishing these are verified by Wikipedia editors. The people who write article in Wikipedia is called "Wikipedians". Wikipedians who write the articles have to give one or more than one (if needed) suitable category names. Categorizing articles is the basic classification rule in Wikipedia. This part of the classification in Wikipedia is done by general people. At the same time this encyclopedia's category structure is used to classify web resources, to link and map with controlled vocabularies, open linked data structure based on Resource Description Framework (RDF) technology and used for document and subject indexing. Wikipedia with its special kind of features is now used for constructing ontologies. This present thesis has the aim to evaluate Wikipedia category system with the help of Dewey Decimal Classification scheme, also known as DDC.

DDC is the most popular library classification scheme. Now a day's most of the library have used DDC in categorising their resources. The present thesis have evaluated Wikipedia category structure in the light of DDC class structure.

Place: Kolkata
Date:

Piyali Ghosh

TABLE OF CONTENTS

Content **Page No.**

LIST OF TABLES

Chapter 5 Findings

Chapter 6 Conclusions

LIST OF FIGURES

Figures Page No.

Chapter 2: Literature Review

Chapter 3 Research Design

Chapter 4 Data Analyses

Chapter 5 Findings

Introduction

Chapter 1: Introduction

Classification pervades all the various activities of our life. Most of us are unaware of the fact that classification is used in every moment of our lives to great extent. After experiencing the outer world from different angels different notions of different things are received by sense organs. Before installing all those first impressions the brain duly classify these so that in future the mind can recall the notion when needed. For these process mind converts unorganized thoughts and impressions into recognizable patterns. All these exercise are completed in very short time that we unaware about it. Basically without classification, human progress would be impossible. That's why from ancient time to this modern age classification acts as a main topic of scientific research. The science of classification or in popular term 'Taxonomy' is being widely used in mathematics, physics, natural sciences, social sciences, and library and information sciences. In this age of information explosion taxonomy helps increasing knowledge and facilitate retrieving of those knowledge in an organised way. In recent years many research works are conducted for inducing taxonomies automatically. In the form of taxonomy, machine readable semantic knowledge would help in information retrieval. But all these can be possible and be easy with the existence structured web contents where the data chunks are semantically networked with each other.

Wikipedia at the age of 18 becomes adult member of information resources which provides multilingual semi-structured content. One of the best thing about Wikipedia is it's Category Network (WCN). WCN refers to the sematic networks which links Wikipedia categories, sub-categories, articles with one another. Now a days Wikipedia therefore is used for creating large scale taxonomies like

WikiTaxonomy (Ponzetto & Strube, 2008), WikiNet (Nastase et al. 2010), YAGO (Suchanek, Kasneci, & Weikum, 2007), BDPedia (Auer et al., 2007), etc. When Wikipedia is being used to classify web content, it is very natural to find answer of a question how Wikipedia itself organise its content or what is the technology or process used for.

1.1. The aim of this study

Wikipedia is a very dynamic and quickly growing encyclopedia which is written in a specific way. Not only that the structure of this huge knowledgebase is also managed and organized in a specific way. In this encyclopedia the vast collection of articles on different topics are arranged through interlinked categories and subcategories. The structure of this knowledgebase is maintained and managed by a category system. The proposed study will evaluate this category structure of Wikipedia. This evaluation is necessary because in this age of information explosion Wikipedia becomes a knowledge hub. In this people encyclopedia the article writing, editing, category-sub category assignment to the written articles all these are done by the writer himself. Evaluation is needed to know and to explore the classification system and the arrangement patterns of pages exits in this vast mass generated encyclopedia. Evaluation is also needed to visualize the category structure exists in this encyclopedia.

1.2. Necessary concepts required for the study

In this present study Wikipedia category structure would have been evaluated in the light of Dewey Decimal Classification System (DDC). Therefore concepts on library classification, DDC, Wikipedia, Wikipedia category structure etc. are required to proceed to this study.

1.2.1. Concepts of library classification

Generally classification means grouping, ordering, and systematic organizing of the objects. According to Mann classification “is the arranging of things according to likeness and unlikeness. It is the sorting and grouping of things, but, in addition, classification of books is a knowledge classification with adjustments made necessary by the physical form of books.” (Kumar, 2010).

Library classification is the library system of arrangement adopted in library to carry out systematically placing or arranging library materials. In 21st century science and technology developed a lot and changes were taking place which brought drastic changes in the library classification as well. With a development of Information and Technology, library has become automated, modern, advanced and developed. Since it is became modern and computerized the library have become knowledge resource centre. In library classification mostly deals with documents and main target is to arrange these in a permanent order, for providing the require materials and information for the readers at the right time. The purpose of the web classification is giving right information at the right time to the users.

1.2.1.1.Dewey Decimal Classification System

The Dewey Decimal Classification (DDC) was conceived by Melvil Dewey in 1873 and first published in 1876 by OCLC. It is a general knowledge organization tool that is continuously revised to keep pace with knowledge. The DDC is accessed through Web Dewey, a frequently updated subscription service maintained by OCLC (“Dewey summaries”, 2024).

Conceptual framework

At the broadest level, the DDC is divided into ten main classes, which together cover the entire world of knowledge. Each main class is further divided into ten divisions, and each division into ten sections (not all the numbers for the divisions and sections have been used). The main structure of the DDC is presented in the DDC Summaries. **The first summary** contains the **ten main classes**. **The second summary** contains the **hundred divisions**. The **third summary** contains the **thousand sections**. The headings associated with the numbers in the summaries have been edited for browsing purposes, and do not necessarily match the complete headings found in the schedules. DDC includes followings as 10 main classes:

Table 1: DDC 10 Main Classes

Notations	**Main Classes**
000	Computer science, information & general works
100	Philosophy & psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography

The DDC is the most widely used classification system in the world. Libraries in more than 138 countries use the DDC to organize and provide access to their collections, and DDC numbers are featured in the national bibliographies of more than sixty countries. The DDC has been translated into over thirty languages.

1.2.2. About Wikipedia

Wikipedia is a free Internet encyclopedia that allows its users to edit almost any article accessible. Wikipedia is the largest and most popular general reference work on the Internet and is ranked among the ten most popular websites. Wikipedia is owned by the non-profit organization Wikimedia Foundation ("Wikipedia", 2024). It has now played a role of a primary research tool of college students. At the same time, faculty and researchers increasingly turn to the site. After introducing students to Wikipedia, teachers might begin by using the site as an entry point into deeper and more creative research than typical assignments require (Crovitz, & Smoot, 2007). As in conventional encyclopaedias, Wikipedia's English language encyclopedia contains articles about topics traditionally considered important, but it also includes entries on all manner of contemporary popular culture and current events. New articles are created daily, existing articles are improved by thousands of volunteer contributors, known as "Wikipedians" to reflect the most up-to-date and accurate knowledge. Its constantly evolving nature has allowed Wikipedia to function in ways that print-based, expert-written reference sources cannot.

1.2.2.1 Features of Wikipedia

Wikipedia seeks to create a summary of all human knowledge in the form of an online encyclopedia, with each topic covered encyclopaedically in one article. Wikipedia website provides following information about their project.

- Wikipedia puts control into the hands of users, who decide what topics are covered and at what depth.
- It is mentioned in the website that each article in Wikipedia must be about a specific topic and the article should be encyclopaedic and it should not a dictionary entry or dictionary-like.
- The topic of the articles must have been covered in mainstream media or major academic journal sources that are independent of the article's subject.
- Further, Wikipedia intends to convey only knowledge that is already established and recognized. It must not present original research.
- For this encyclopedia, the readers, not the encyclopedia, are ultimately responsible for checking the truthfulness of the articles and making their own interpretations.
- Wikipedia must not take sides. All opinions and viewpoints, if attributable to external sources, must enjoy an appropriate share of coverage within an article. This is known as neutral point of view (NPOV). NPOV is a fundamental principle of Wikipedia and of other Wikimedia projects. It says that, all encyclopaedic content on Wikipedia must be written from a neutral point of view, representing fairly, proportionately, and as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic.
- Wikipedia's initial anarchy integrated democratic and hierarchical elements over time. An article is not considered to be owned by its creator or any other editor and is not vetted by any recognized authority.

- Wikipedians may dispute, for example by repeatedly making opposite changes to an article.
- Each article and each user of Wikipedia has an associated "Talk" page. These form the primary communication channel for editors to discuss, coordinate and debate. Wikipedia gives guideline to its writer for writing articles in this mass encyclopedia. Before writing anything in it people may practice according to the guideline in the “Sandbox” available in Wikipedia (Ghosh & Mondal, 2018).

1.2.2.2 Use of Wikipedia for different purpose

Wikipedia is the encyclopedia of everything for everyone. With all controversies regarding its accuracy and authoritativeness, this encyclopedia has been used for following purposes:

- The Parliament of Canada’s website referrers to Wikipedia's article on same-sex marriage in the "related links" section of its "further reading" list for the Civil Marriage Act (“Civil Marriage Act”, 2017);
- Supreme Court of India in a judgment, while defining "Common Law Marriage", cites Wikipedia's article as a source (Legally India, 2010);
- US federal courts and the World Intellectual Property Organization for their supporting information uses Wikipedia (Arias, 2007);
- Content appearing on Wikipedia has also been cited as a source and referenced in some US intelligence agency reports (Aftergood, 2007)
- The journal *RNA Biology" and the RNA family database (Rfam) consortium make a project in their collaboration which requires a summarise page in

Wikipedia for future use and then peer review the page before publishing it in Wikipedia (Butler, 2008);

- On September 16, 2007, The Washington Post newspaper reported that Wikipedia had become focal point in the 2008 US election campaign (Vargas, 2007);
- Law students of Monash University Australia have been assigned to write Wikipedia articles as an exercise in clear and succinct writing for an uninitiated audience (Witzleb, 2009);
- Over 400 universities in the United States and Canada have used Wikipedia in the classroom. At California State University System some faculty members create interesting and unique assignments which involve Wikipedia or Wikipedia like work (Soules, 2015);
- Xiao and Askin (2012) examined the popularity of Wikipedia as a venue for academic publishing. To do these authors had compared the process of publishing a peer reviewed article in Wikipedia and the open access journal model. According to authors, Wikipedia reduces the overhead cost for publication, makes possible the post- publication correction.
- Library Wikipedia collaboration becomes the new paradigm of library services (Ghosh & Mondal, 2018). Wikipedia is a powerful tool for public libraries, especially in parts of the world where access to other reference sources is expensive. The categories like culture, geography, health, etc. includes in this encyclopedia provide huge number of articles in different places which can fulfill the various information needs of users (Ghosh & Mondal, 2018).

1.2.2.3. Statistics related Wikipedia

Wikipedia is a multilingual encyclopedia with is rank 5th according to Semrush (List of most visited websites, 2024). There are currently 332 language editions of Wikipedia. They are called language versions, or simply Wikipedias. Thirteen of these Wikipedias have over one million articles each (English, Swedish, Cebuano, German, Dutch, French, Russian, Waray-Waray, Italian, Spanish, Polish, Vietnamese and Japanese). Four more have over 500,000 articles (Portuguese, Chinese, Ukrainian and Catalan), 42 more have over 100,000 articles, and 75 more have over 10,000 articles. The largest, the English Wikipedia, has over 61,391,032 articles as of September, 2024. Total categories are around 20 lacks. Registered users (list of members) or Wikipedians are 47,927,027 ("Wikipedia: statistics", 2024).

1.3. Organization of thesis

The thesis is divided into six individual chapters. This present chapter is **Introduction** (Chapter 1). This chapter includes necessary concepts on DDC and Wikipedia and the aim of the study.

Chapter 2 is dedicated for **Literature Review**. Thematically all literatures have been arranged and discussed. This chapter also includes observations on the reviewed literature.

Chapter 3 is **Research Design**. This chapter includes research problems, significance of the problems, research questions, sampling methods, data collection process, and the methodology.

Chapter 4 is for representing **Data Analyses**. On two variables data has been analysed: Resemblance and Influence. The analysed data has been represented through tables, charts.

Chapter 5 is **Findings.** The findings on two variables Resemblance and Influence are described through tables, pie charts, column charts.

Chapter 6 is **Discussion** which includes recommendations, scope for further research and conclusions.

Then the thesis is accompanied by necessary references.

Literature Review

Chapter 2: Literature Review

Wikipedia systematically organize its huge collection of articles in different categories and subcategories. But in this open source Mediawiki project, anyone can write and edit articles, anyone can assign categories and subcategories to the written articles. Many studies were conducted on Wikipedia category structure. 'The evaluation of Wikipedia's category structure' is the research problem fixed for this present study. Mainly the proposed study will evaluate how Wikipedia classify its content with its category system. Therefore the focus of searching was given on those literature which are related to the evaluation of Wikipedia categories. This chapter presents a systematic review on the category structure in Wikipedia.

2.1. Methodology followed

Review of literature can be represented chronologically or thematically. In this study, the literature are reviewed thematically. A good review should be comprehensive with all authentic and relevant sources; selective which are appropriate to reveal the research work done on a given issue; synthesis of key points associated with the given issue; balanced different ideas (Steward, 2004). With the help of literature review one can find the literary gap exists in a specific field. The present study reviews the collected literature systematically under different themes. Articles were collected from different databases, like, ERIC, Google Scholar, Emerald Insight, NDLTD, Shodhganga, ALIS, etc. The keywords used for searching were 'Wikipedia category', 'Classification of articles in Wikipedia', 'usage of Wikipedia category', 'Wikipedia category assignment', etc.

2.2. Scope of the literature

Researchers of various fields are now engaged in Wikipedia research. Paul and Pramanik (2013) compared Wikipedia with Encyclopedia Britannica. Though Encyclopedia Britannica is an authentic, reliable and renowned encyclopedia but the popularity of Wikipedia cannot be ignored. In another study, Wikipedia's scientific information was compared with Encyclopedia Britannica. The work was published in Nature (Giles, 2005). Many other research works on Wikipedia are available, conducted from different point of views. Some of these are the content analysis of its articles on different subjects, like philosophy (Ghosh & Mondal, 2016), political science (Brown, 2011), medical science (Hasty, et al, 2013; Clauson, 2008) library and information science (Ghosh & Mondal, 2018), etc., evaluation of its content accuracy (Ford & Sen, 2013; Read, 2006; Rector, 2008), faculty perception (Soules, 2015) and librarian perception (Luyt, 2010) of Wikipedia, Wikipedia's topical coverage (Halavais & Lackaff, 2008), semantic coverage of its relational statements (Holloway, Bozicevic & Borner, 2007), and many more.

But the focus of searching was given on those literature which were related to Wikipedia category evaluation. Then the collected articles were filtered with the help of three main themes, like Works on evaluation of Wikipedia as a knowledge organization tool, Wikipedia category structure, and comparison between Wikipedia and classification tools.

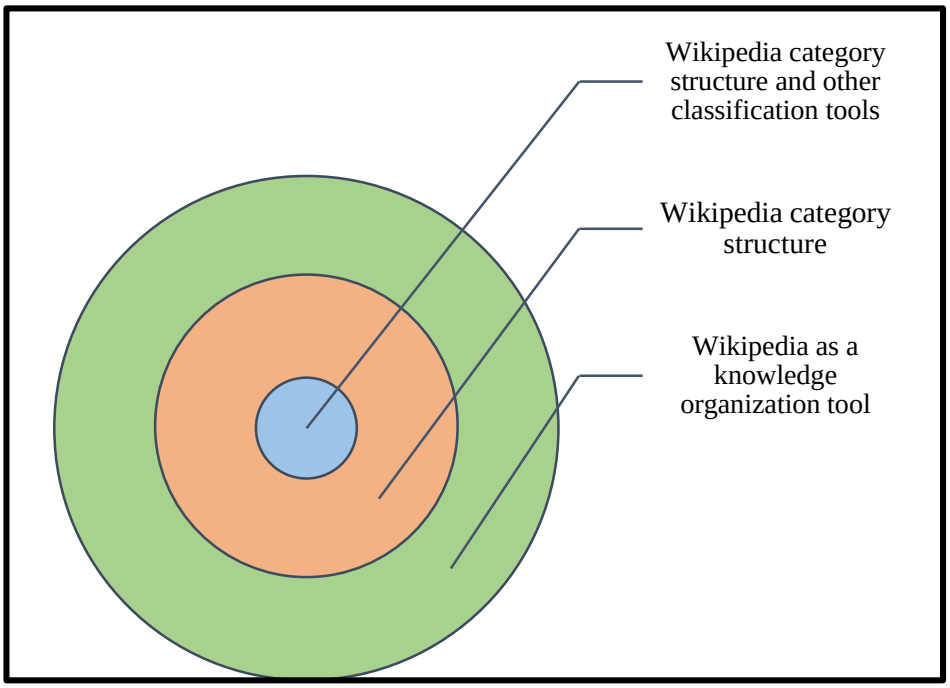


Figure 1: Thematic distributions of collected literature

The above figure demonstrate the way how this present study would review all those works from “Wikipedia as a knowledge organization tool” to “Wikipedia category structure” and then to “Wikipedia category structure and other classification tools”.

2.3. The reviewed literature

Reviewing literature is the main part of any research works. Literature review should be done from neutral point of view from which researcher may conclude about the literary gap found in a given topic. The resent part of this chapter will represent all the relevant literature with this motto. All the literature are represented thematically one after one.

2.3.1. Literature on “Wikipedia as a knowledge organization tool”

It can be easily understood by the above literature that Wikipedia’s category structure is now become a focus of many research works. Wikipedia has been

scrutinised from different point of views. It is the right time to represent those literature which revealed that Wikipedia is being or may be used as knowledge organization tool. Jakob Voss (2009) defined the ways how Wikipedia can be used for that purpose. Wikipedia provides a term with its definition through the article on that term and also in different parallel language. Moreover semantic linking with different articles and hierarchy of categories (concepts) make it a knowledge organization tool that can be used for subject indexing and for many other purposes. Because of these special features Wikipedia now is started to link and map with other controlled vocabularies, open linked data structure based on Resource Description Framework (RDF) technology. Wikipedia's category network may be used as a tool for document indexing (Chahine et al., 2011). Its category structure may be defined as taxonomy of different concepts. For the purpose of document indexing, the terms selected from those documents will have to be connected with categories and subcategories provided in Wikipedia. Domain specific thesaurus is the highly useful resource for us. Now a days researchers (Milne, Medelyan & Witten, 2006) are now trying to build such kind of domain thesaurus with terms and links which can be mined automatically from Wikipedia. Actually in Wikipedia all the pages of specific names depict some specific concepts and in each page there are suitable links which express hierarchical, associative, equivalence relationships. These features can be suitably fitted with the features of thesaurus and of course a thesaurus of specific domain. Wikipedia is now used to extract name entities and synonyms from its content to make extended name entity dictionary (Higashinaka et al, 2012). Wikipedia with its special kind of features is now used for constructing ontologies. There are exist different Wikipedia based ontologies. Most popular ontology is DBpedia (Lehmann et al., 2014). DBpedia's content is based on the information found in Wikipedia, especially from

Infoboxes, between language links and redirects. Another two are Yago (Suchanek et al., 2008; Hoffart et al., 2013) and Catriple (Liu et al., 2008). Yago uses Wikipedia category structure and integrate it with WordNet (Miller, 1995). Catriple also uses Wikipedia category structure. It is based on the category names found in Wikipedia and extracts the facts about the concepts found in the articles under those category names.

2.3.2. Literature on "Wikipedia category structure"

Wikipedia's category structure was introduced in 2004. After then many research works were conducted on Wikipedia category. An evolution of Wikipedia category structure by Suchecki, Salah and Gao (2012) focused on the clustering of articles exist in Wikipedia defined by its category system. The work investigated the category structure of English Wikipedia from 2004 to 2008. Wikipedia category system is not purely hierarchical. Here the similar categories are inter-related. To know about inter-relation between categories a research work (Szymanskki & Duch, 2012) designed visualization map. Then the map was used to know the type of relation found in categories. The map also helped to reach the category from its similar category. Some research work concentrated on the semantic relations found in Wikipedia categories.

The category found in Wikipedia is a system for classifying the whole content of the encyclopedia. Here users classify the content and give names which they found appropriate to the articles. Pablo and Monti (2016) researched on the folksonomy associated with Wikipedia category system and the impact of the classification system on Wikipedia categories. One year later a research work (Cai et al, 2017) found that Wikipedia category system due to its local classification has low precision and recall.

Researchers built automatic classification system where Wikipedia has no need for human effort. They used inbuilt info boxes of the articles. But using automatic classifier for Wikipedia articles may become useless due to improper classification of articles. Therefore they can be miscategorised. Colgrove, Neidert and Chakoumakos (2011) found the fault in automatic classification used for Wikipedia content.

Later on research works were going on with different other kind of topics. Wikipedia categories can be used for different purposes, such as for predicting person's age and gender (Kosgi et al., 2014), for entity ranking (Kaptein & Kamps, 2013), or as a source of knowledge acquisition (Nastase & Strube, 2008). Many research works revealed that Wikipedia Category structure are now used for document classification. It may be used as a tool to identify document topic (Schonhofen, 2012). The work found that Wikipedia category structure on a given topic represents an ontology and taxonomy of all the related category names, whereas in another work (Pappu, 2009) Wikipedia categories were used to classify named entities found in the taxonomy of in it. This classification was then used further to extract semantic relations among those name entities collecting from Wikipedia articles. Wikipedia in now considered as the reliable source for scientific information (Giles, 2005). Eyal (2018) described how scientific information are organised in Hebrew Wikipedia into a proper discipline.

2.3.3. Literature on "Wikipedia category structure and other classification tools"

Wikipedia organises its huge number of articles is a specific way where these articles are arranged under a suitable category and sub category. Links between categories and sub categories, categories to its articles make Wikipedia now a very

special reference source. Now Wikipedia is being used for categorising web videos (Chen et al., 2010), different web resources (Medeiros et al., 2018), learning resources (Meyer, Rensing, & Steinmetz, 2018). Wikipedia can also be utilised for improving document classification (Wang et al., 2009) for classifying short text of Tweet or any micro blogging sites (Pratapa & Kumar, 2014), classifying reputation dimension of a particular politician, singer or company reflecting in Tweet (Quresi et al., 2014). In this recent trend of research Wikipedia is now critically investigated as a classification tool. As a result it is very natural to move towards a question, i.e., is there any similarity between Wikipedia and other traditional classification tools? Or what makes Wikipedia different from any classification tool? To find answer of these questions literature search were conducted. It was found that there are very limited works regarding these issues. A research team of Knowledge Space Lab (KSL), the Virtual Knowledge Studio of the 26 Royal Netherlands Academy of the Arts and Sciences worked on this issue. KSL was a project begun in 2009 with the goal of creating a knowledge map of the evolution of science by tracking the evolution of knowledge in the then emergent Wikipedia (Scharnhorst et al., 2016). To reach the goal researchers, namely Salah, Gao, Suchecki and Scharnhorst (2010; 2011) used Universal Decimal Classification (UDC). The team compared the category structure of UDC with Wikipedia. Wikipedia is ever emerging encyclopedia depend on crowd sourcing. The category and subcategory assignment to the article also is done by the users themselves. This work basically mapped the emerging structure of category links of Wikipedia (which is created socially) to the UDC class structure (creating formally using classification theory).

2.4. Observations

At the time of searching literature it was fixed that Wikipedia category system will be evaluated. The collected literature shows that there are very limited research works which deal with the comparison of Wikipedia category structure with any existing classification scheme. After four years of its publishing Wikipedia introduced category system to classify its content and to assign a suitable category and subcategory name its article. Another point is that Wikipedia is now using to classify other web contents. Therefore Wikipedia category structure should be examined with any existing most popular classification scheme. The above literature review shows a lack in this regard. More research work should be done on this issue where Wikipedia's category structure could be scrutinized on the basis of any classification scheme.

Research Design

Chapter 3: Research Design

3.1. Statement of the problem

The problem of the proposed research may be stated as: **The evaluation of Wikipedia's category structure**

3.2. Significance of the problem

The present study has the problem to evaluate Wikipedia's category structure and it is very significant. Because in this age of information explosion Wikipedia becomes a knowledge hub. In this people encyclopedia the article writing, editing, category-sub category assignment to the written articles all these are done by the writer himself. Evaluation is needed to know and to explore the classification system and the arrangement patterns of pages exits in this vast mass generated encyclopedia. Evaluation is also needed to visualize the category structure exists in this encyclopedia.

3.3. Research questions

The following research questions are designed. These questions will direct this research work to find out solution of previously stated problem, i.e., "Evaluation of Wikipedia Category Structure".

i) Is there any similarity between the category structure found in Wikipedia with the structure of DDC?

In Wikipedia the category structure means the tree like structure of categories where all the sub-categories placed under parent categories and the sub-categories under any

parent category contain many other sub-sub-categories. This present study will compare this structure with DDC where the universe of knowledge is first divided into ten main classes and then the ten classes are subdivided into hundred divisions and the hundred divisions are further divided into thousands sections.

ii) Is the classification of Wikipedia categories into one or more sub-categories or sub-sub categories influenced by DDC classification system?

In Wikipedia, all categories are classified on the basis of their contents or topics and every category therefore is assigned one or more than one sub category names. Is there any resemblance between this kind of classification system exist in Wikipedia with DDC classification system? The above question is formulated to find out the answer.

3.4. Methodology

Wikipedia to organize its huge knowledgebase uses its own classification system. Here all the articles are arranged under categories and subcategories. The purpose of this paper is to evaluate Wikipedia's classification system lying into its category structure. To fulfill this purpose Wikipedia category structure will be compared with the structural hierarchy of DDC.

Why DDC?

DDC is chosen as because DDC and Wikipedia both are American projects. DDC is the most popular library classification scheme widely used all over world in different types of libraries. Like multilingual Wikipedia, DDC has been translated into over thirty languages. DDC is the first modern classification scheme for libraries, it is

regularly being revised and adopted according to entry of new subjects and giving right place in the schedules. DDC came in 1876. It is efficiently working not only in print, but in online environment also. Currently, OCLC offers DDC 23rd edition in web version 'WebDewey 2.0' to the users. It has completed a long journey of more than 100 years from its first edition in 1876 to the 23rd edition published in 2011. The abridged 15th edition was published in February 2012. DDC is based on a discipline to be organized with fields, into sub-fields and topics. Likewise Wikipedia has organised its collection in parent categories, into sub categories, sub sub categories and so on.

Before going to describe the methodology it will be convenient to give a brief description on the category structure of Wikipedia and the hierarchical structure of DDC.

Hierarchical structure of DDC

At the top level, the DDC is divided into ten main classes, which together cover the entire world of knowledge. The first summary contains these ten classes. Each main class is further divided into ten divisions. The second summary contains these hundred divisions. Each division is divided into ten sections. The third summary contains the thousand sections. As a system of library classification the DDC is arranged by discipline, not by subject.

Wikipedia Category Structure

Wikipedia has tree like category structure formed by creating links between inter-related categories. Category: Main Topic Classification is a list of Wikipedia's major topic classifications. They are used throughout Wikipedia to organize the presentation of links to articles in its various reference systems, including Wikipedia's overviews,

outlines, glossaries, lists, portals, indices, and categories (“Category: main topic classifications”, 2019). The purpose of this category is to group major topic classifications in one place. Any main category may contain or "branch into" subcategories, and the subcategories are further divided into some other categories (“Wikipedia: categorization”, 2024). The articles or pages are simultaneously placed into the categories or sub categories.

Methodology used for the study

The present study is a qualitative research. Qualitative research involves collecting and analysing non-numerical / qualitative data (e.g., text, video, or audio) to understand concepts, opinions, or experiences. It can be used to gather in-depth insights into a problem or generate new ideas for research. Qualitative research is commonly used in the humanities and social sciences, in subjects such as anthropology, sociology, education, health sciences, history, etc. (Bhandari, 2023).

Qualitative research may be in the form of content analysis, thematic analysis, textual analysis, discourse analysis. In this present study the web content of Wikipedia (category structure of Wikipedia) is to be evaluated in the light of DDC classification scheme. Therefore this study would be a qualitative content analysis. For qualitative research method purposive sampling is more appropriate and widely used.

3.5. Scope of the study

Wikipedia is available in total 332 languages. Among them English Wikipedia has the highest number of articles and categories. As on September, 2024, English Wikipedia includes total number of 61,391,032 articles and around 20 lacks categories.

("Wikipedia: Statistics", 2024). The present study have evaluated the categories of **English Wikipedia.** The categories are evaluated in the light of **DDC's 23rd** edition on three main classes, namely **Philosophy, Social Sciences, and Science.**

3.6. Sampling

For this study purposive sampling is conducted and "Philosophy", "Social Sciences", "Science" categories of Wikipedia are chosen. The category structure of these three disciplines are to be evaluated with the DDC class 100 (Philosophy), class 300 (Social science), class 500 (science).

1. Now in this regard one question may be appeared. Why only philosophy, social science, science? The academic disciplines are divided as Arts and Humanities, Social Sciences, Science. From these, Social sciences and science are taken. From Arts and Humanities only Philosophy is selected, as it is a mother of all subjects.

2. Another thing is, in my dissertation research guided by Late. Associate Prof. Tarun Kumar Mondal Sir, it was found that Wikipedia is a best reference source on "Indian Philosophy" in comparison with Stanford Encyclopedia of Philosophy (SEP) and Internet Encyclopedia of Philosophy (IEP). That study have recommended a research work on the whole subject "philosophy" and its availability on Wikipedia (Ghosh & Mondal, 2018).

3.7. The whole research design

The next table shows whole research design including research questions, variables, methodologies and units of measurements in DDC and in Wikipedia.

Table 2: Research Design

SN	Research Questions	Variables	Methodology	Units of measurements in DDC	Units of measurements in Wikipedia
1.	Is there any similarity between the categories structure found in Wikipedia with the structure of DDC?	Resemblance	One to one mapping of the section names of the each divisions of the main classes "philosophy", "social sciences", "science" of DDC with Wikipedia categories	Section names of the each divisions of the main classes "philosophy", "social sciences", "science"	Similar name of the DDC sections of the each divisions of the main classes "philosophy", "social sciences", "science"
2.	Is the classification of Wikipedia categories into one or more categories influenced by DDC classification system?	Influence	Comparing the parent categories of the Wikipedia category with DDC	DDC divisions of the main classes "philosophy", "social sciences", "science"	Wikipedia categories with the same name of DDC divisions of the main classes "Philosophy", "Social Sciences", "Science"

3.8. Data collection process

From DDC the divisions of three main classes were noted down. The classes are "Philosophy", "Social Science", and "Science". The next two charts display the way how data has been collected.

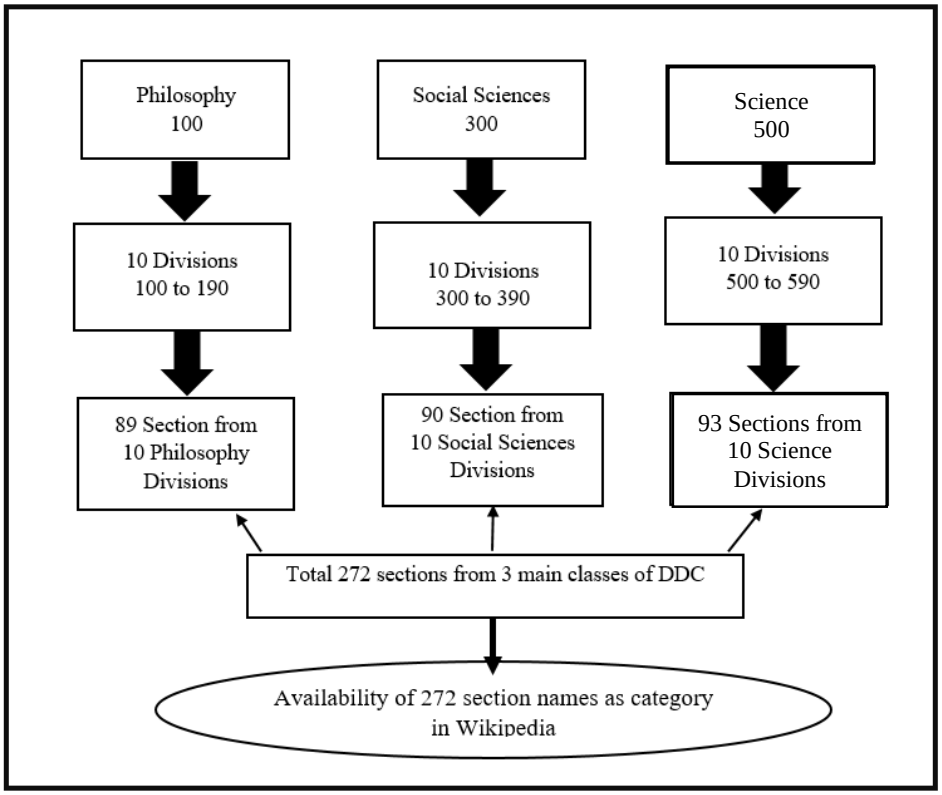


Figure 2: Data Collection Process for the variable Resemblance

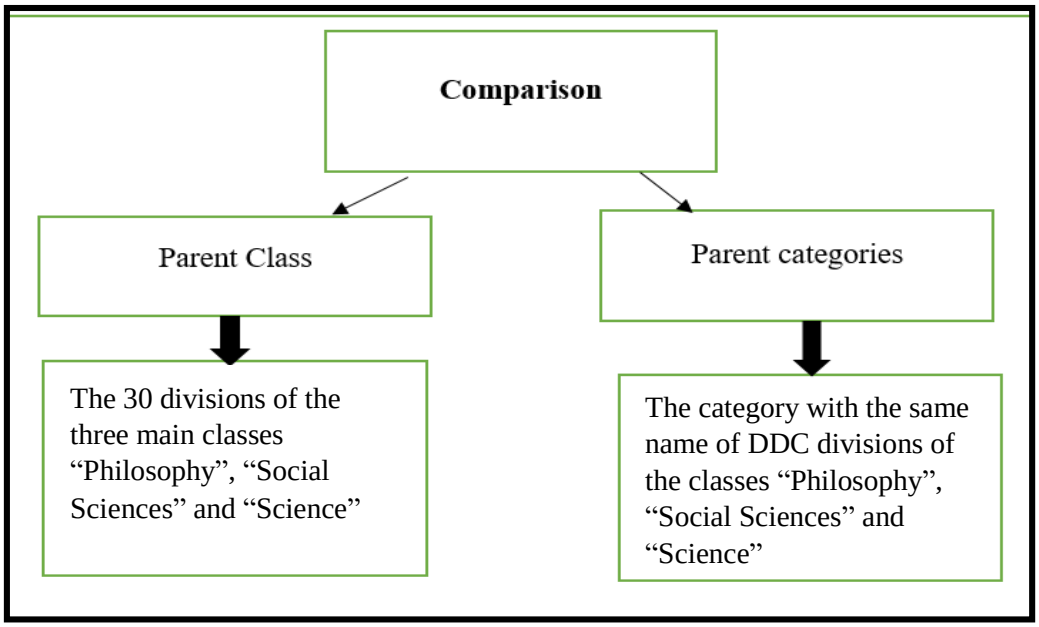


Figure 3: Data Collection process for the variable Influence

Data Analyses

Chapter 4: Data Analyses

In this part of the research work the data collected from DDC and Wikipedia has been analysed. On the basis of following variables the collected data are analysed.

- Resemblance
- Influence

Therefore this chapter is subdivided into further two sections under the name of the above variables.

4.1. Data analyses on 'Resemblance'

'Resemblance' means similarity. This part of the research is find out the availability of Wikipedia category similar to / resemblance with DDC section names of the main classes philosophy, social sciences and science. Methodology taken was one to one mapping of the section names of the each divisions of the main classes philosophy, social sciences, science of DDC with Wikipedia categories. Section 4.1 is subdivided into three sub-sections: Resemblance for the main class Philosophy (section 4.1.1), Resemblance for the main class Social Sciences (section 4.1.2), Resemblance for the main class Science (section 4.1.3).

4.1.1. Resemblance for the main class Philosophy (class 100)

The following table shows the data coding plan of the variable "Resemblance". Before going to see the analyses on "Resemblance" variable it is convenience to understand the data coding method.

Table 3: Data coding for "Resemblance Variable"

Data	Coded Data
Compete resemblance	Yes
Partly resemblance	Yes (partly)
No resemblance	No
Not Assigned	NA

The collected data is presented through this following table. The table represent the availability of resemblance with the word "Yes", the non-availability of resemblance with the word "No", the partially resemblance with the word "Yes (partly)". For the unassigned sections the word used is "NA", i.e., not applicable.

Table 4: Availability of Wikipedia category similar to DDC "Philosophy" section

DDC Divisions	DDC Sections	Availability of Wikipedia category
Philosophy 100	Philosophy and Psychology (100)	Yes
	Theory of philosophy (101)	No
	Miscellany (102)	No
	Dictionaries and encyclopaedias (103)	Yes
	Unassigned class (104)	NA
	Serial publication (105)	No
	Organizations and management (106)	Yes
	Education, research and related topics (107)	Yes
	Kinds of persons treatments (108)	No

	Historical and collected persons treatment (109)	No
Metaphysics 110	Metaphysics (110)	Yes
	Ontology (111)	Yes
	Unassigned class (112)	NA
	Cosmology (113)	Yes
	Space (114)	Yes
	Time (115)	Yes
	Change (116)	Yes
	Structure (117)	Yes
	Force and energy (118)	Yes
	Number and quantity (119)	Yes (partly)
Epistemology 120	Epistemology, causation & humankind (120)	Yes
	Epistemology, theory of Knowledge (121)	Yes
	Causation (122)	Yes
	Determinism and Indeterminism (123)	Yes
	Teleology (124)	Yes
	Unassigned class (125)	NA
	The self, Consciousness of person, personal identity, personality, personhood (126)	Yes (partly)
	The unconscious and subconscious (127)	Yes (partly)
	Humankind, philosophical anthropology, philosophy of human life, Psychology of human life (128)	Yes (partly)
	Origin and Density of Individual Soul, immortality, incarnation, reincarnation, transmigration (129)	Yes (partly)
	Parapsychology & occultism (130)	No

Parapsychology and Occultism 130	Parapsychological and occult methods for achieving well-being, happiness, success (131)	No
	Unassigned class (132)	NA
	Specific topics in parapsychology and occultism (133)	No
	Unassigned class (134)	NA
	Dreams and Mysteries (135)	Yes (partly)
	Unassigned class (136)	NA
	Divinatory graphology (137)	No
	Physiognomy (138)	Yes
	Phrenology (139)	Yes
Specific Philosophical Schools and Viewpoints 140	Specific philosophical schools (140)	Yes
	Idealism and related systems (141)	Yes
	Critical Philosophy (142)	Yes
	Bergsonism and Intuitionism (143)	Yes (Partly)
	Humanism (144)	Yes
	Sensationalism (145)	No
	Naturalism (146)	Yes
	Pantheism (147)	Yes
	Dogmatism, Eclecticism, liberalism, Syncretism & traditionalism (148)	Yes
	Other philosophical systems (149)	Yes
Psychology 150	Psychology (150)	Yes
	Unassigned class (151)	NA
	Perception, movement, emotions & drives (152)	Yes (partly)
	Conscious Mental Processes and intelligence (153)	No
	Subconscious & altered states (154)	No
	Differential and Developmental Psychology (155)	Yes (partly)
	Comparative Psychology (156)	No

	Unassigned class (157)	NA
	Applied Psychology (158)	Yes
	Unassigned class (159)	NA
Philosophical Logic 160	Logic (160)	Yes
	Induction (161)	Yes
	Deduction (162)	Yes
	Unassigned class (163)	NA
	Unassigned class (164)	NA
	Fallacies and Sources of Error (165)	Yes (partly)
	Syllogisms (166)	Yes
	Hypotheses (167)	Yes
	Argument and persuasion (168)	Yes
	Analogy (169)	Yes
Ethics 170	Ethics (170)	Yes
	Ethical System (171)	Yes
	Political ethics (172)	No
	Ethics of family relationships (173)	No
	Occupational ethics (174)	Yes
	Ethics of recreation, leisure, public performances, communications (175)	No
	Ethics of Sex and Reproduction (176)	Yes
	Ethics of Social Relations (177)	Yes
	Ethics of consumption (178)	No
	Other ethical norms (179)	No
Ancient, Medieval, Eastern Philosophy 180	Ancient, Medieval, Eastern Philosophy (180)	No
	Eastern Philosophy (181)	Yes
	Pre-Socratic Greek Philosophies (182)	Yes (Partly)
	Sophistic, Socratic, Related Greek Philosophies (183)	Yes
	Platonic Philosophy (184)	Yes
	Aristotelian Philosophy (185)	Yes

	Skeptic and Neoplatonic philosophies (186)	Yes
	Epicurean Philosophy (187)	Yes
	Stoic Philosophy (188)	Yes
	Medieval Western Philosophy (189)	Yes
Modern Western and Other Non-eastern Philosophy 190	Modern western philosophy (190)	No
	Philosophy of United States and Canada (191)	Yes
	Philosophy of British Isles (192)	Yes
	Philosophy of Germany and Austria (193)	Yes
	Philosophy of France (194)	Yes
	Philosophy of Italy (195)	Yes
	Philosophy of Spain and Portugal (196)	Yes
	Philosophy of Russia (197)	Yes
	Philosophy of Scandinavia and Finland (198)	Yes
	Philosophy in other geographic areas (199)	Yes

Out of 100 sections, total number of 89 sections from DDC main class philosophy are assigned, 11 sections are unassigned. After word by word searching in Wikipedia category tree it's found that total 69 categories are found in Wikipedia, out of which 58 categories are completely resemblanced with DDC sections and 11 categories are partly resemblanced. 20 sections are not found in Wikipedia as category.

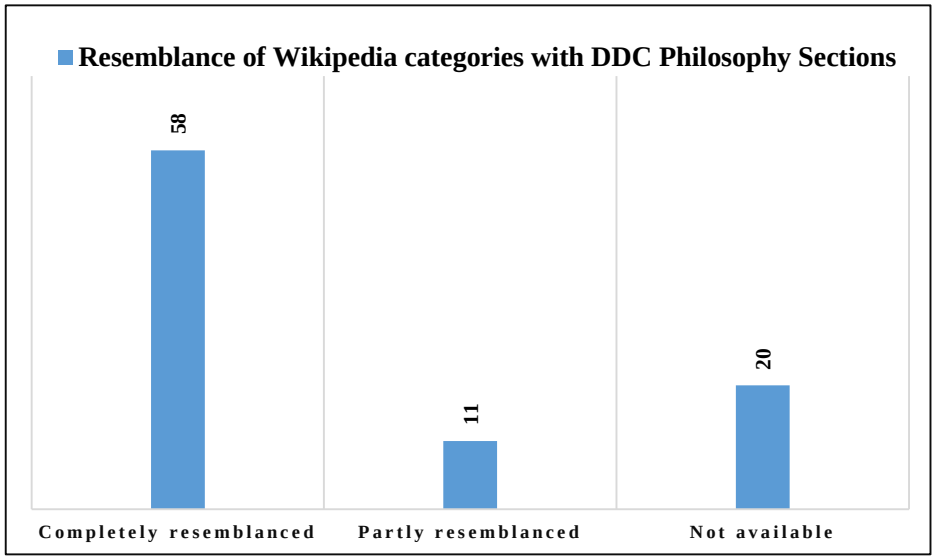


Figure 4: Resemblance of Wikipedia categories with DDC Philosophy sections

The column chart represents the availability of resemblance of Wikipedia category with 89 DDC philosophy sections. This overall data division wise is further represented in next few pages.

Division wise data representation for the main class Philosophy

1. Resemblance for the division Philosophy (100)

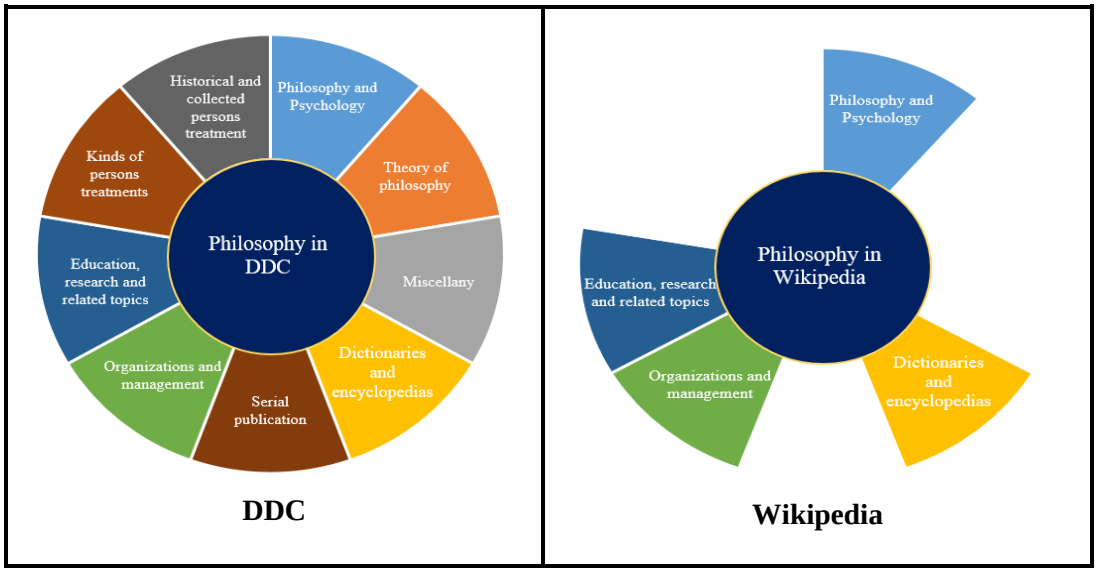


Figure 5: Resemblance for the DDC division Philosophy

1. Resemblance for the division Metaphysics (110)

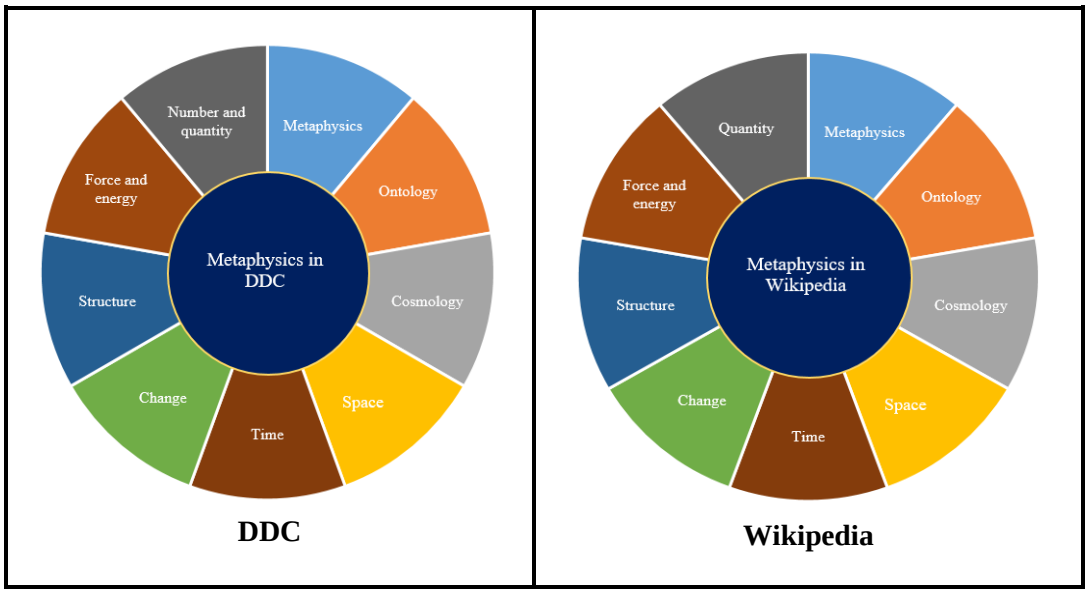


Figure 6: Resemblance for the DDC division Metaphysics

2. Resemblance for the division Epistemology (120)

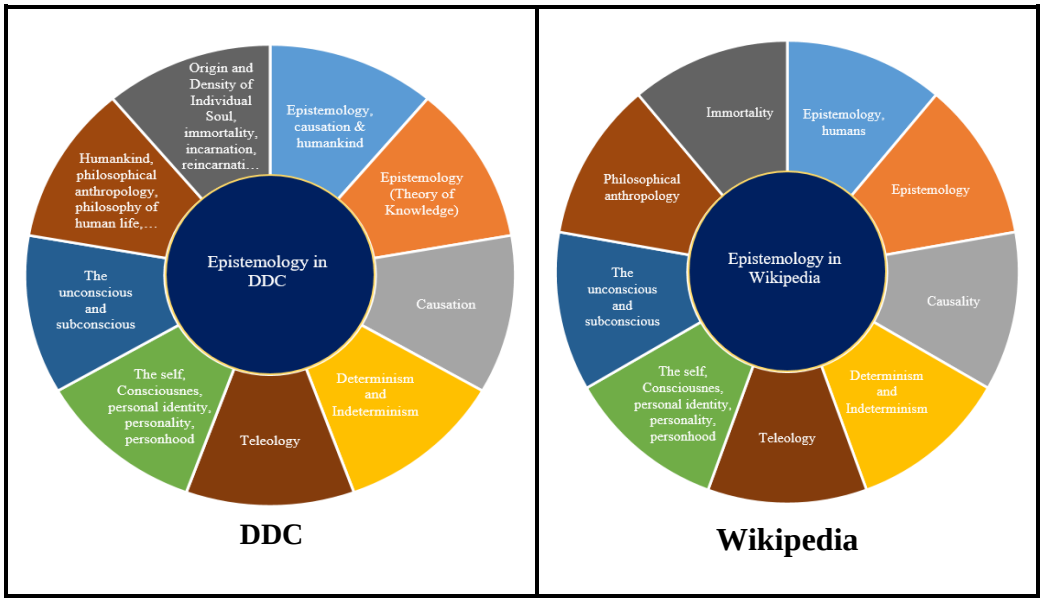


Figure 7: Resemblance for the DDC division Epistemology

3. Resemblance for the division Parapsychology and Occultism (130)

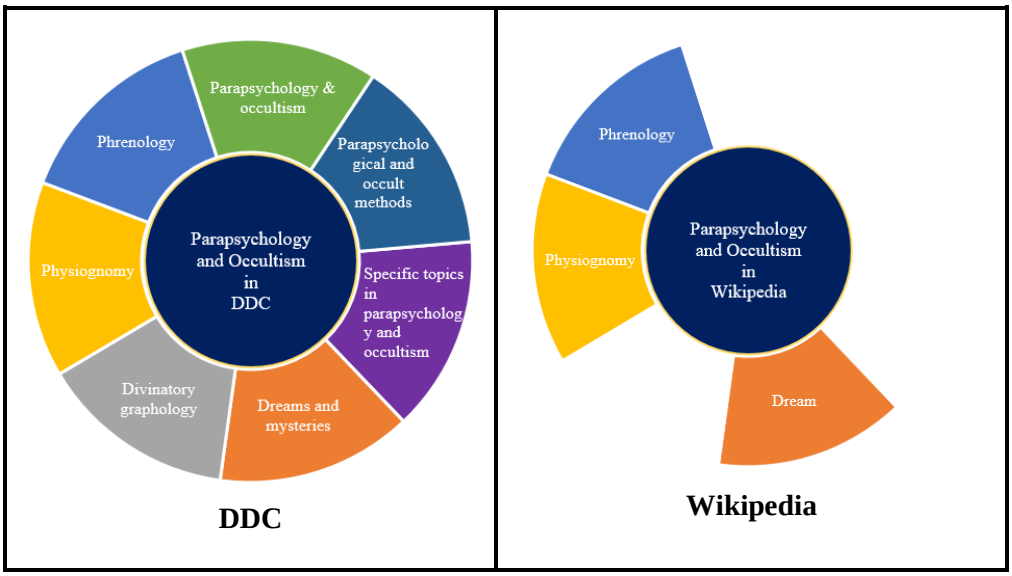


Figure 8: Resemblance for the DDC division Parapsychology and Occultism

4. Resemblance for the division Specific Philosophical Schools and Viewpoints (140)

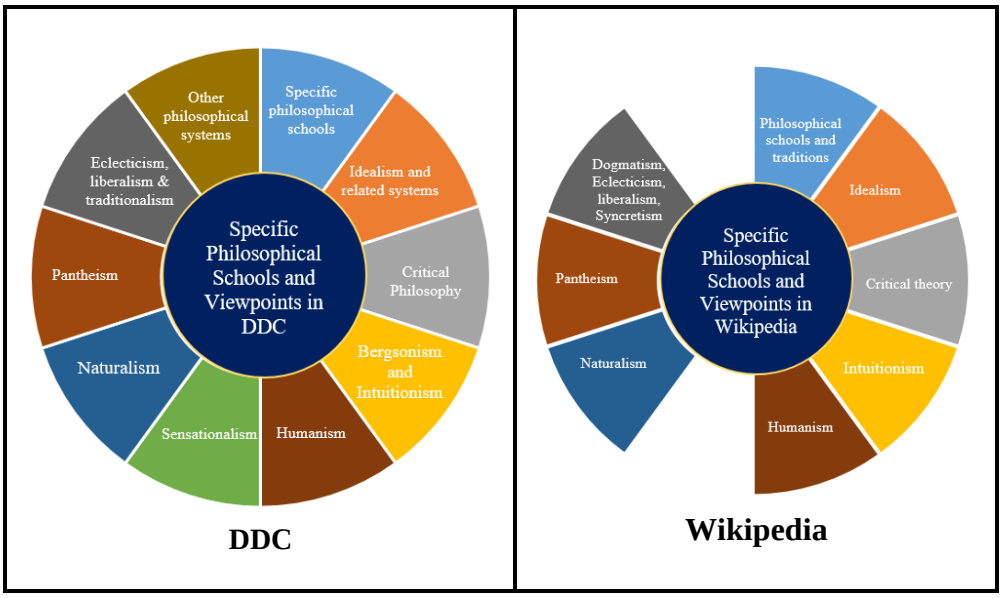


Figure 9: Resemblance for the DDC division Specific Philosophical Schools and Viewpoints

5. Resemblance for the division Psychology (150)

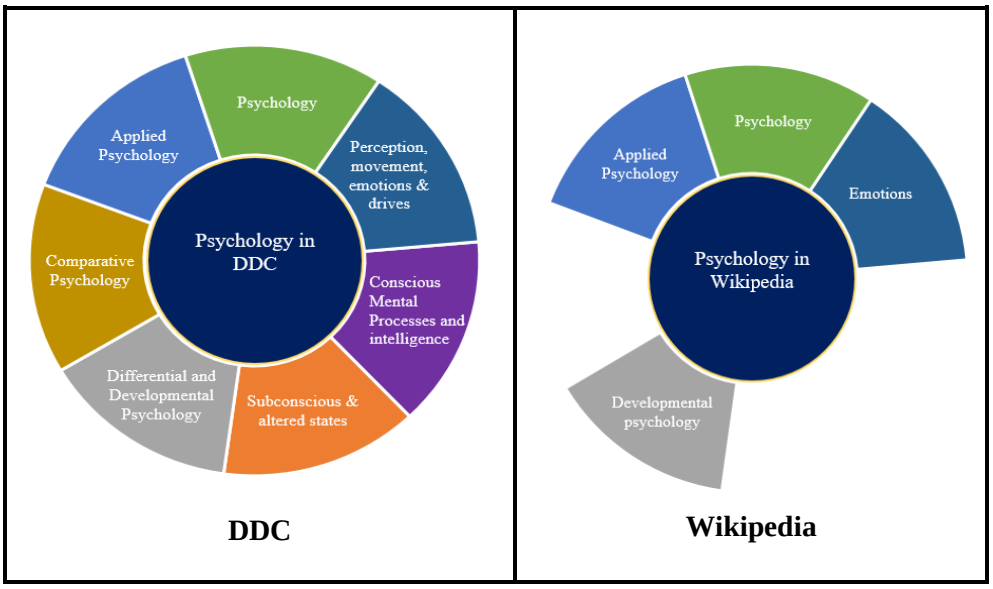


Figure 10: Resemblance for the DDC division Psychology

6. Resemblance for the division Philosophical Logic (160)

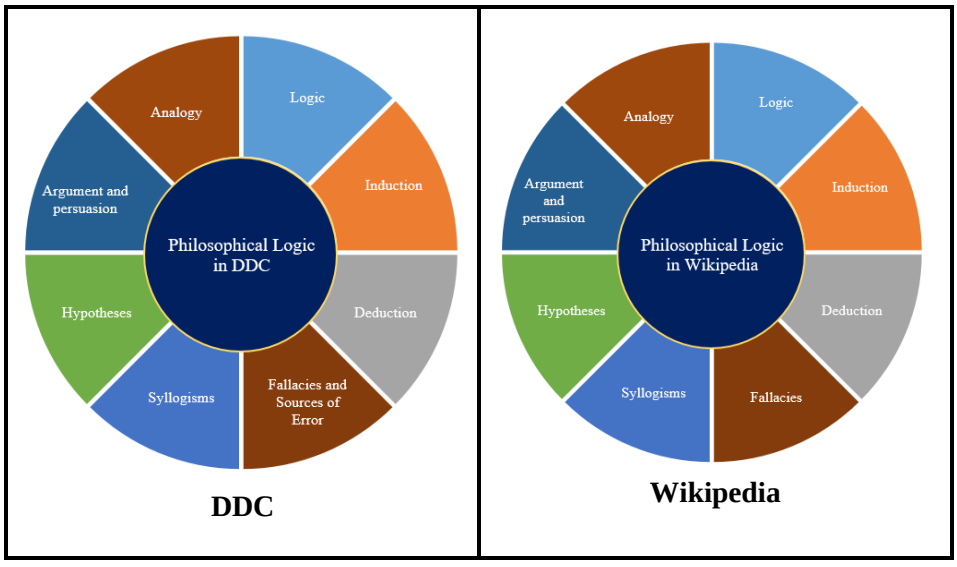


Figure 11: Resemblance for the DDC division Philosophical Logic

7. Resemblance for the division Ethics (170)

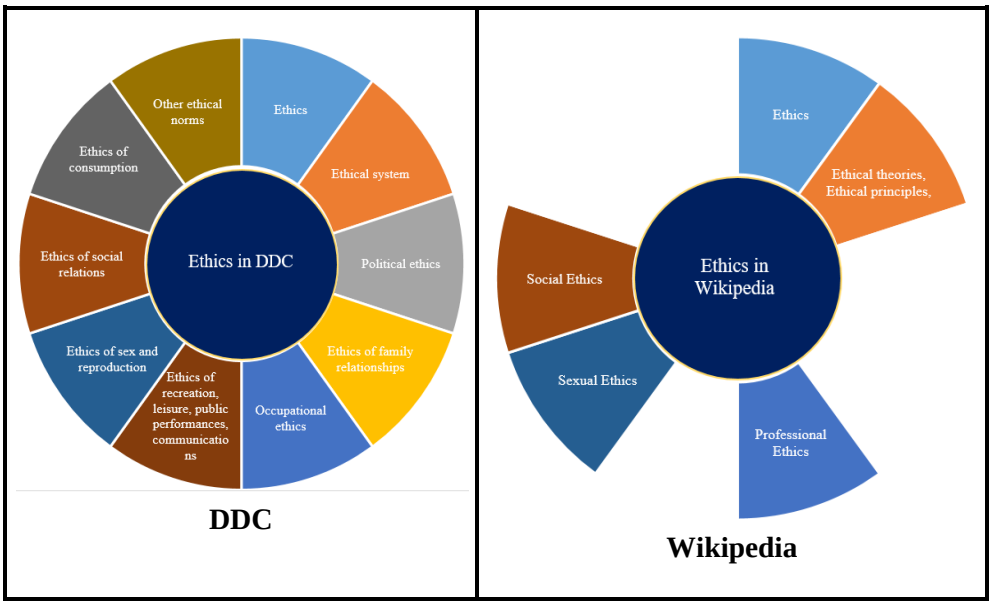


Figure 12: Resemblance for the DDC division Ethics

8. Resemblance for the division Ancient, Medieval, Eastern Philosophy (180)

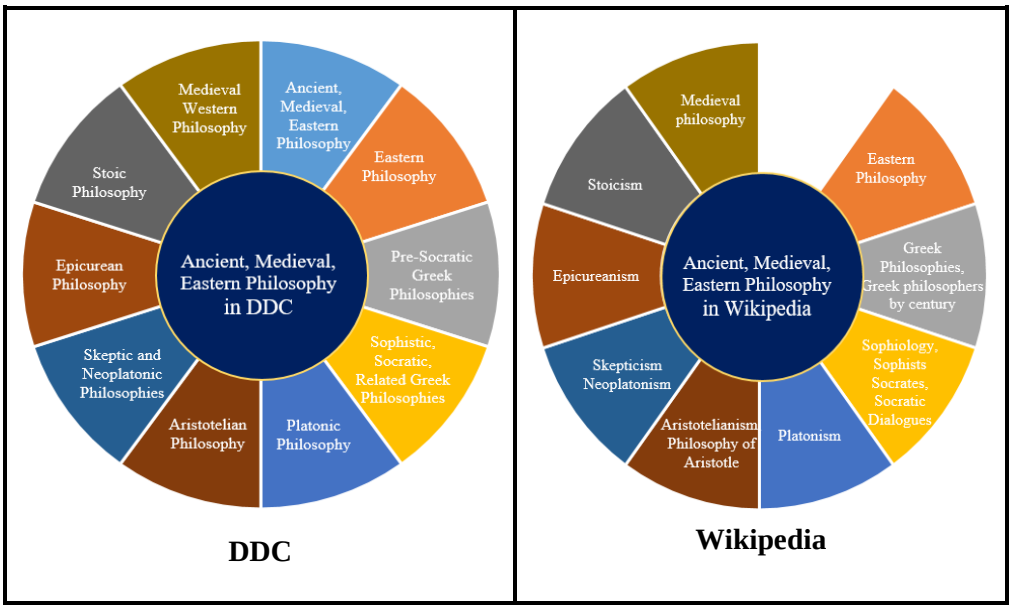


Figure 13: Resemblance for the DDC division Ancient, Medieval, Eastern Philosophy

9. Resemblance for the division Modern Western and Other Non-eastern Philosophy (190)

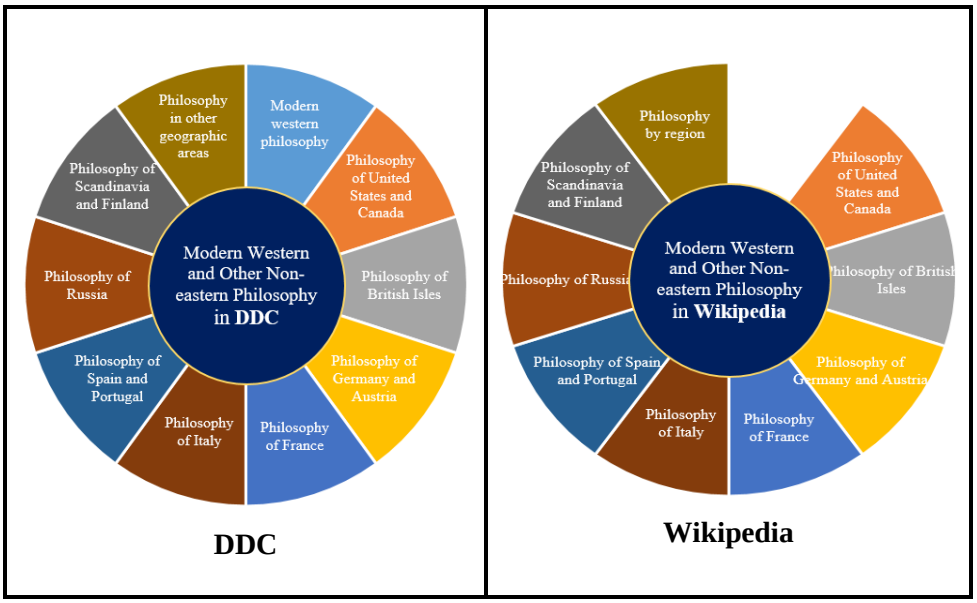


Figure 14: Resemblance for the DDC division Modern Western and Other Non-eastern Philosophy

The next column chart represents the status of availability of categories in Wikipedia with the same name of DDC divisions of the main class philosophy. The chart represents the availability or non-availability of Wikipedia categories as a whole.

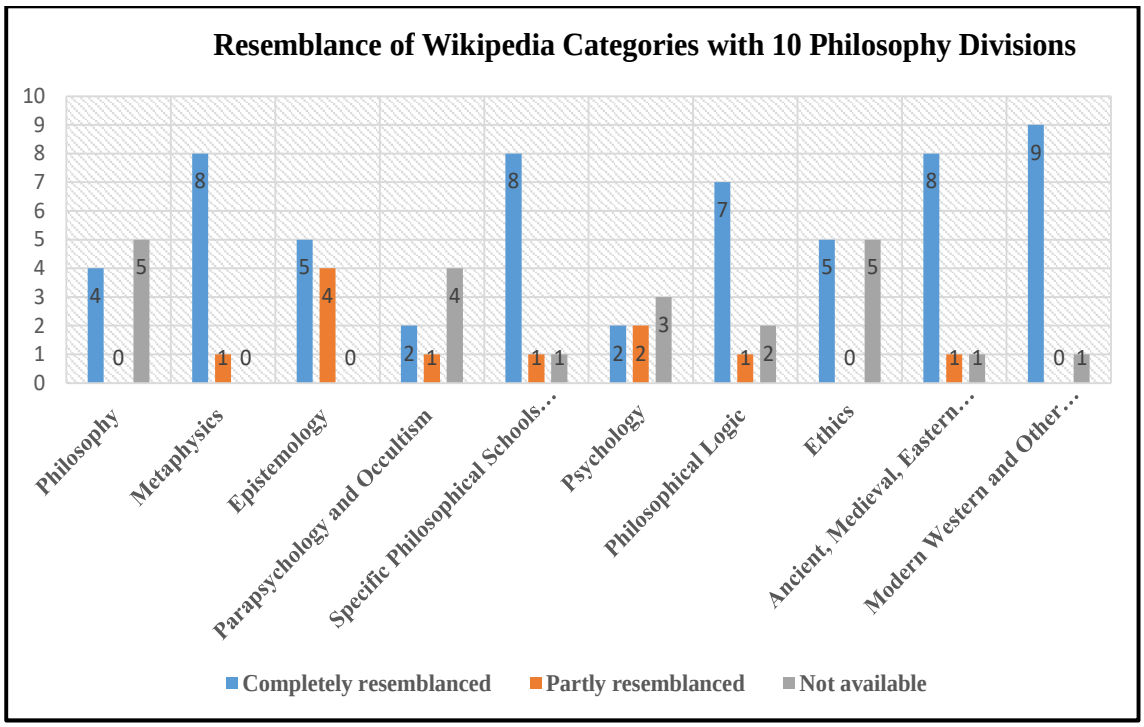


Figure 15: Resemblance of Wikipedia Category with 10 Philosophy divisions

4.1.2. Resemblance for the main class Social Sciences (class 300)

The collected data is presented through this following table. The table represent the availability of resemblance with the word "Yes", the non-availability of resemblance with the word "No", the partially resemblance with the word "Yes (partly)". For the unassigned sections the word used is "NA", i.e., not applicable.

Table 5: Availability of Wikipedia category similar to DDC "Social Sciences" section

DDC Divisions	DDC Sections	Availability of Wikipedia category
Social Sciences 300	Social sciences (300)	Yes
	Sociology and anthropology (301)	Yes
	Social interaction (302)	No

	Social processes (303)	No
	Factors affecting social behaviour (304)	No
	Social groups (305)	Yes
	Culture and institutions (306)	Yes (Partly)
	Communities (307)	Yes
	Unassigned Class (308)	**NA**
	Unassigned Class (309)	**NA**
Collections of General Statistics 310	Collections of General Statistics (310)	Yes
	Unassigned Class (311)	NA
	Unassigned Class (312)	NA
	Unassigned Class (313)	NA
	General Statistics of Europe (314)	No
	General Statistics of Asia (315)	No
	General Statistics of Africa (316)	No
	General Statistics of North America (317)	No
	General Statistics of South America (318)	No
	General Statistics of Australia, Pacific Ocean island, Atlantic Ocean island, Arctic island, Antarctica (319)	No
Political Science(Politics and government) 320	Political Science (320)	Yes
	Systems of Governments and States (321)	Yes (Partly)
	Relation of the state to organized groups and their members (322)	No
	Civil and political rights (323)	Yes (Partly)
	The political process (324)	No
	International migration and colonization (325)	Yes (Partly)
	Slavery and emancipation (326)	Yes
	International relations (327)	Yes
	The Legislative Process (328)	No
	Unassigned (329)	NA
Economics 330	Economics (330)	Yes
	Labour Economics (331)	Yes

	Financial Economics (332)	Yes
	Economics of land and Energy (333)	No
	Cooperatives (334)	Yes
	Socialism and related system (335)	Yes
	Public Finance (336)	Yes
	International economics (337)	Yes
	Production (338)	No
	Macroeconomics and related topics (339)	Yes
Law 340	Law (340)	Yes
	Law of Nations (341)	Yes
	Constitutional and Administrative Law (342)	Yes
	Military, defense, public property, public finance, tax, commerce (trade), industrial law (343)	Yes (partly)
	Labor, social service, education, cultural law (344)	Yes (partly)
	Criminal law (345)	Yes
	Private law (346)	Yes
	Procedure and courts (347)	Yes (partly)
	Laws, regulations, cases (348)	Yes
	Law of specific jurisdictions, areas, socioeconomic regions, regional intergovernmental organizations (349)	Yes
Public administration and military science 350	Public administration and military science (350)	Yes
	Public Administration (351)	Yes
	General Consideration of public Administration (352)	No
	Specific fields of public administration (353)	No
	Public Administration of Economy and Environment (354)	No
	Military science (355)	Yes
	Foot forces and warfare (356)	Yes (Partly)
	Mounted forces and warfare (357)	No

	Air and specialized forces and warfare; Engineering and related services (358)	Yes (Partly)
	Sea forces and warfare (359)	No
Social problems and services; association 360	Social problems and services; association (360)	No
	Social problem and services (361)	No
	Social problem and services to group of people (362)	No
	Other social problem and services (363)	No
	Criminology (364)	Yes
	Penal and related institutions (365)	Yes (Partly)
	Secret associations and societies (366)	No
	General clubs (367)	No
	Insurance (368)	Yes
	Associations (369)	No
Education 370	Education (370)	Yes
	Schools and their activities; Special Education (371)	Yes (Partly)
	Primary Education (Elementary education) (372)	Yes (Partly)
	Secondary education (373)	Yes
	Adult education (374)	Yes
	Curricula (375)	Yes
	Unassigned Class (376)	NA
	Unassigned Class (377)	NA
	Higher education (Tertiary Education) (378)	Yes
	Public policy issues in education (379)	No
Commerce, communications, transportation 380	Commerce, communications, transportation (380)	Yes
	Commerce (Trade) (381)	Yes
	International commerce (Foreign Trade) (382)	Yes
	Postal communication (383)	Yes
	Communications (384)	Yes
	Railroad transportation (385)	Yes

	Inland waterway and ferry transportation (386)	Yes
	Water, air, space transportation (387)	Yes
	Transportation (388)	Yes
	Metrology and standardization (389)	Yes
Customs, etiquette, folklore 390	Customs, etiquette, folklore (390)	Yes (Partly)
	Costume and personal appearance (391)	No
	Customs of life cycle and domestic life (392)	No
	Death customs (393)	Yes
	General customs (394)	No
	Etiquette (Manners) (395)	Yes
	Unassigned Class (396)	NA
	Unassigned Class (397)	NA
	Folklore (398)	Yes
	Customs of war and diplomacy (399)	No

Out of 100 sections, total number of 90 sections from DDC main class social science are assigned, 10 sections are unassigned. After word by word searching in Wikipedia category tree it's found that total 55 categories are found in Wikipedia, out of which 45 categories are completely resemblanced with DDC sections and 10 categories are partly resemblanced. 35 sections are not found in Wikipedia as category.

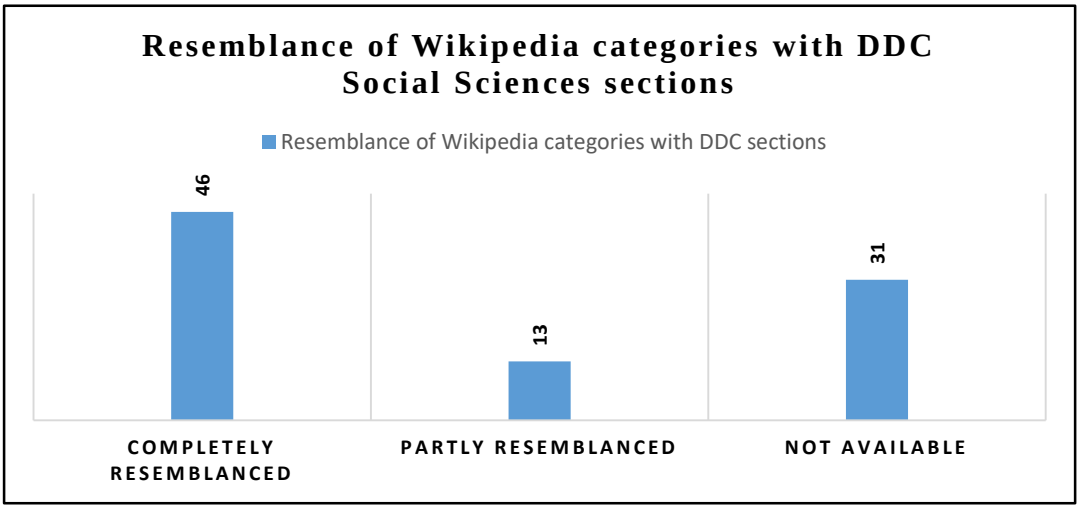


Figure 16: Resemblance of Wikipedia categories with DDC Social Sciences sections

The above column chart represents the availability of resemblance of Wikipedia category with 90 DDC social sciences sections. This overall data division wise is further represented in next few pages.

Division wise data representation for the main class Social Sciences

1. Resemblance for the division Social Sciences (300)

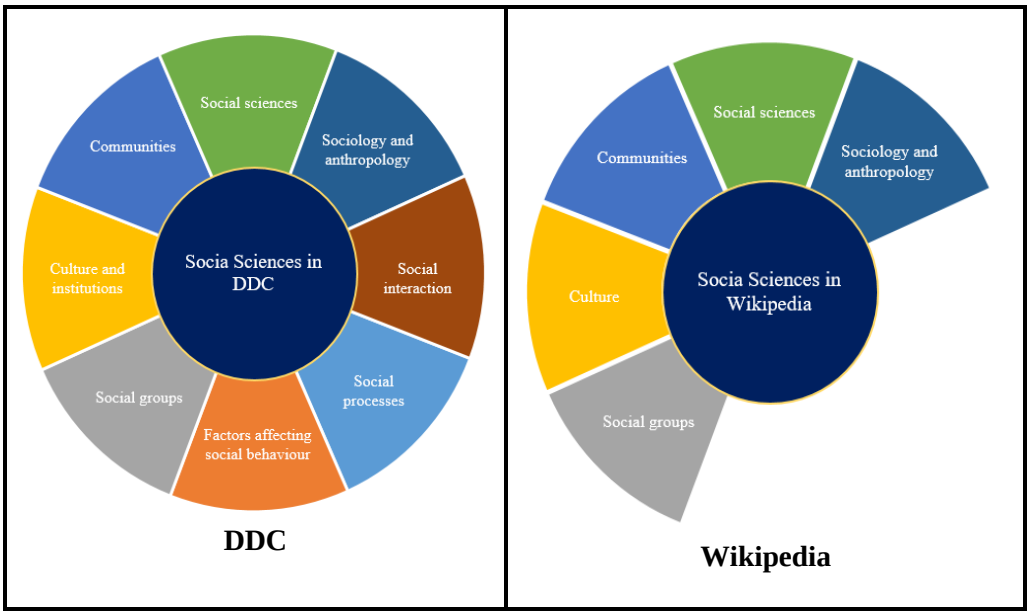


Figure 17: Resemblance for the DDC division Social Sciences

2. Resemblance for the division Collections of General Statistics (310)

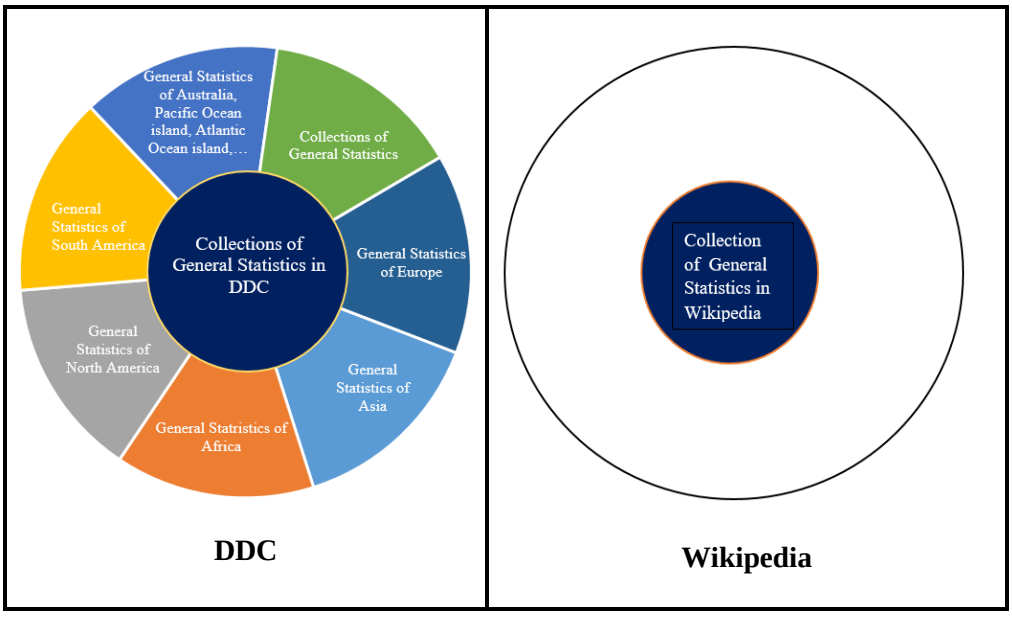


Figure 18: Resemblance for the DDC division Collections of General Statistics

3. Resemblance for the division Political Science (Politics and Government) (320)

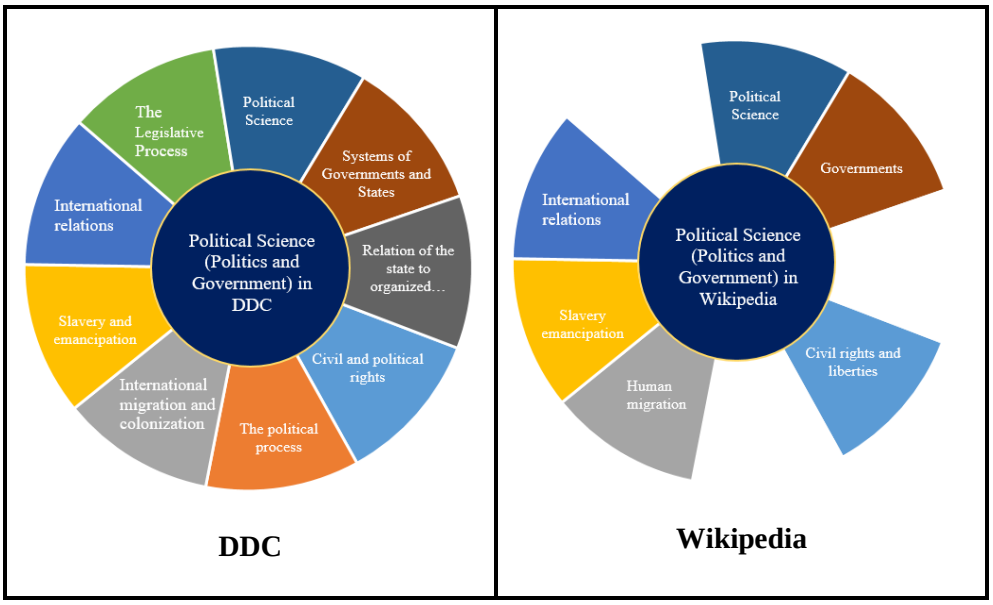


Figure 19: Resemblance for the DDC division Political Science (Politics and Government)

4. Resemblance for the division Economics (330)

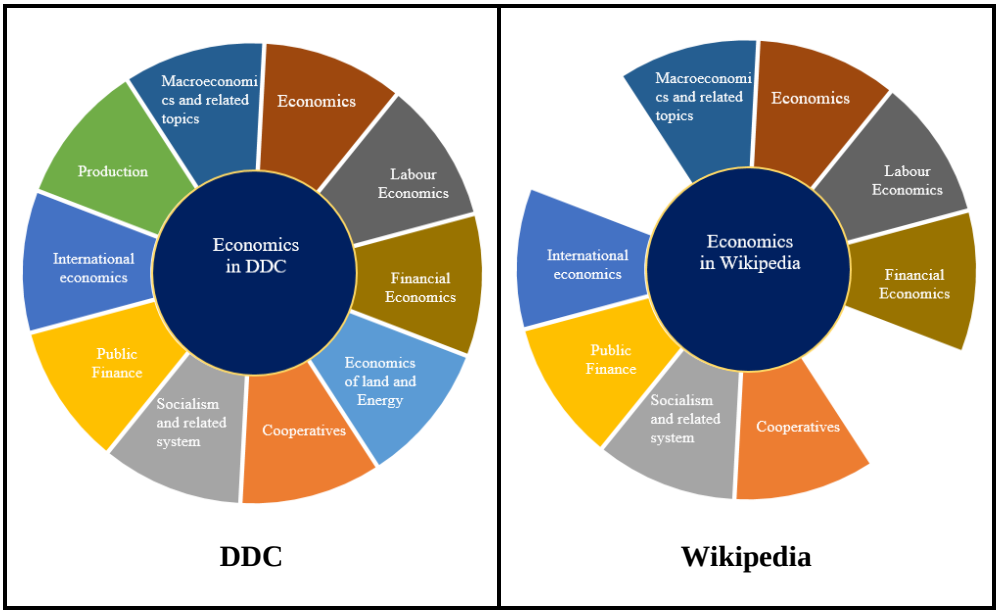


Figure 20: Resemblance for the DDC division Economics

5. Resemblance for the division Law (340)

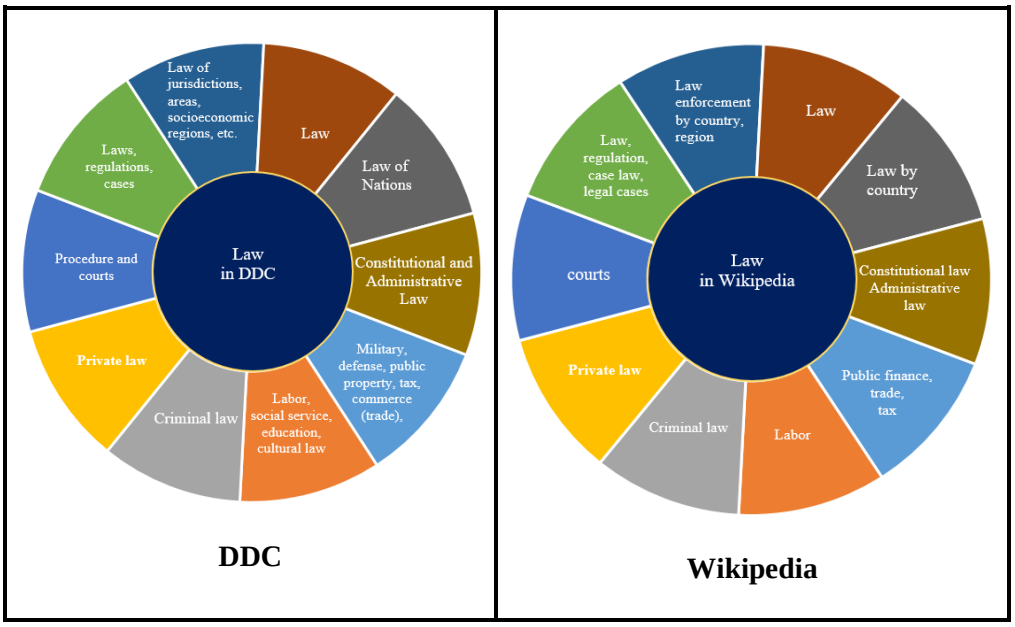


Figure 21: Resemblance for the DDC division Law

6. Resemblance for the division Public Administration and military science (350)

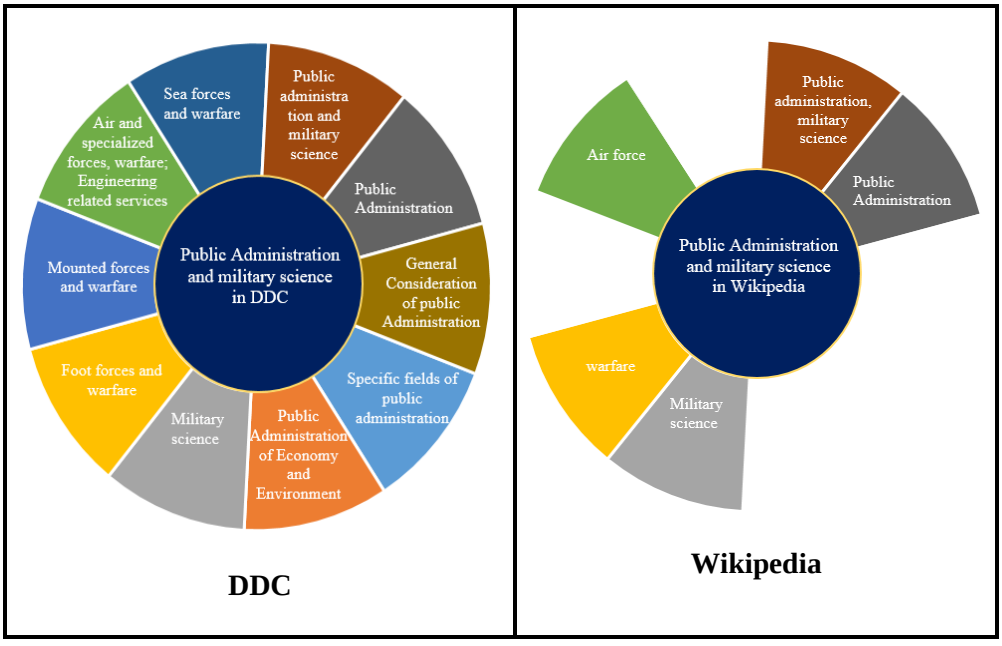


Figure 22: Resemblance for the DDC division Public Administration and military science

7. Resemblance for the division Social problems and services; associations (360)

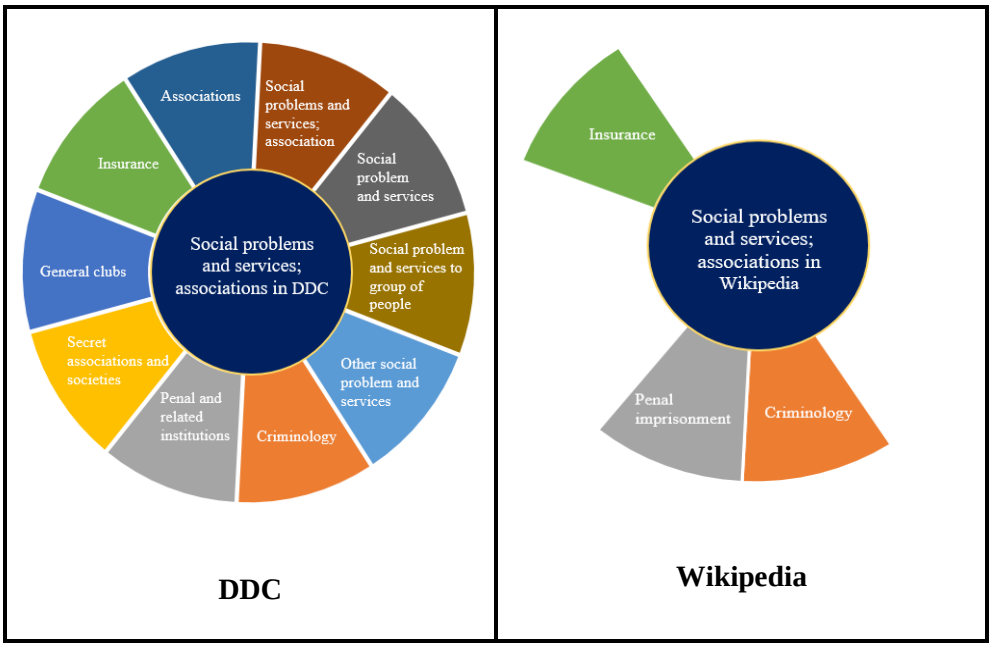


Figure 23: Resemblance for the DDC division Social problems and services; associations

8. Resemblance for the division Education (370)

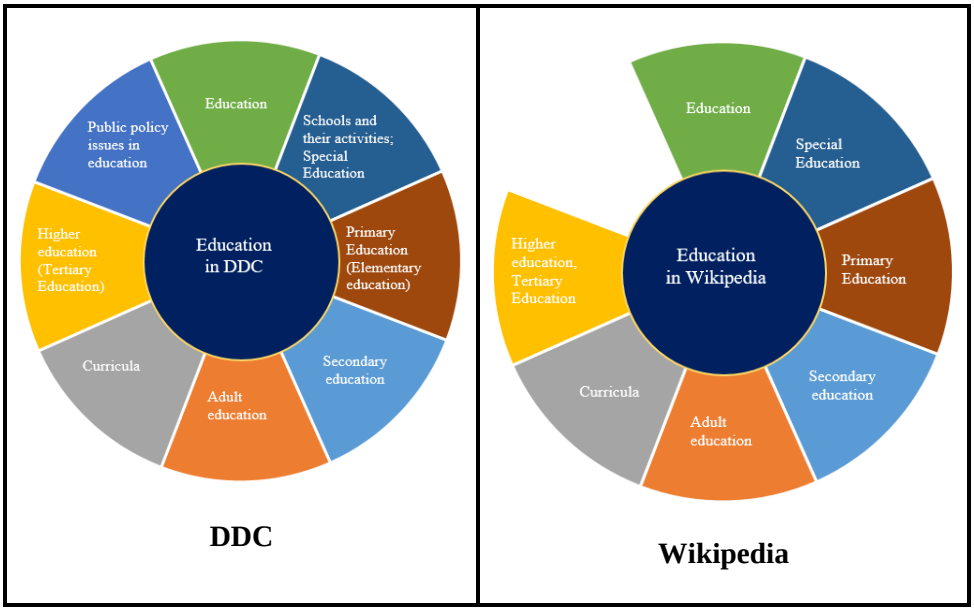


Figure 24: Resemblance for the DDC division Education

9. Resemblance for the division Commerce, communications, transportation (380)

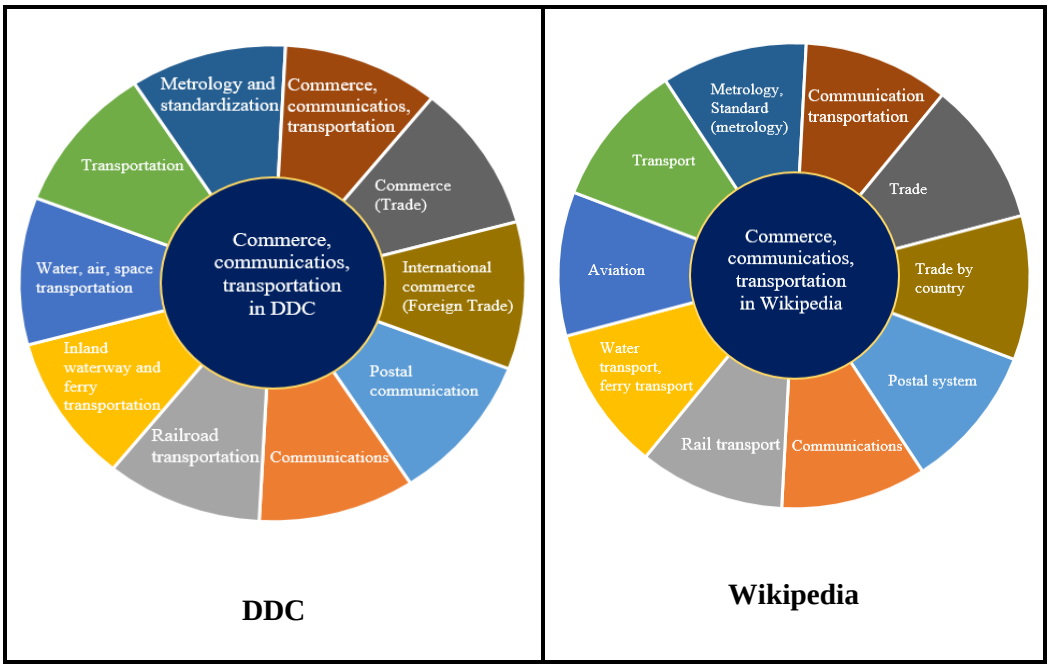


Figure 25: Resemblance for the DDC division Commerce, communications, transportation

10. Resemblance for the division Customs, etiquette, folklore (390)

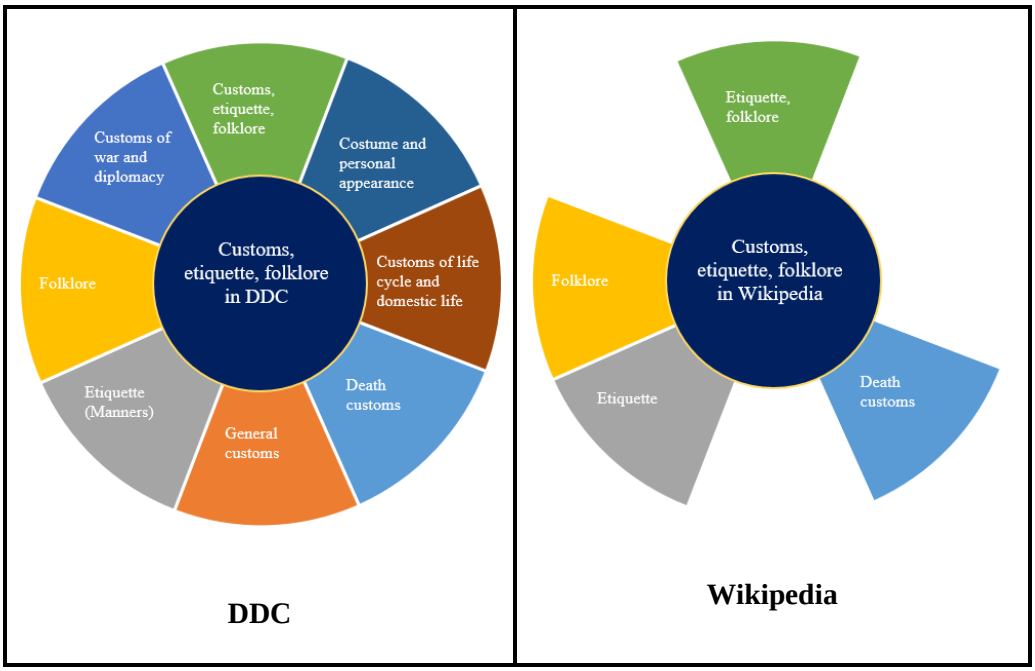


Figure 26: Resemblance for the DDC division Customs, etiquette, folklore

The following column chart represents the status of availability of categories in Wikipedia with the same name of DDC divisions of the main class social sciences. The chart represents the availability or non-availability in whole.

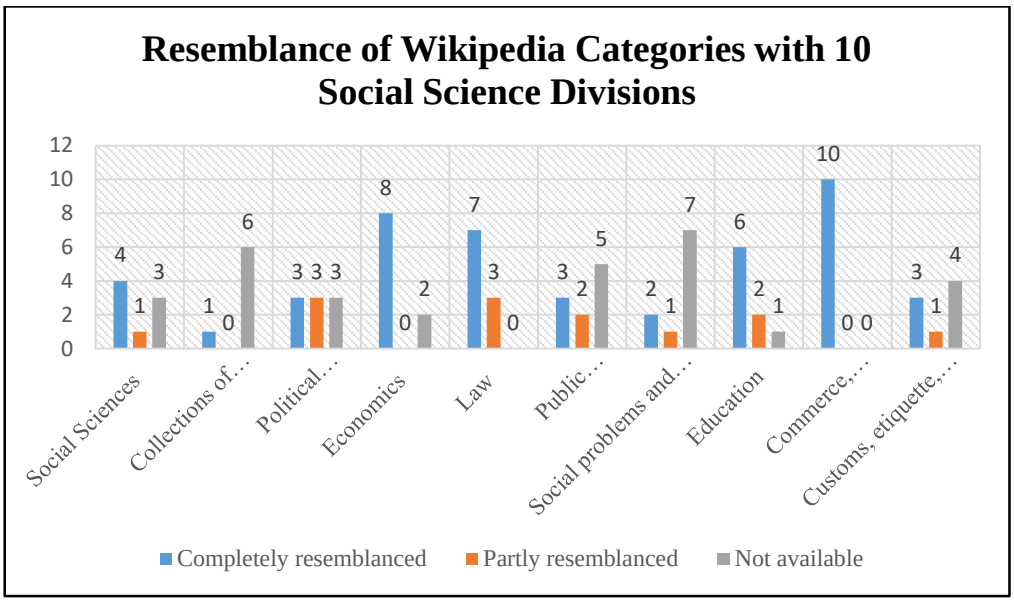


Figure 27: Resemblance of Wikipedia Category with 10 Social Sciences divisions

4.1.3. Resemblance for the main class Science (class 500)

The collected data is presented through this following table. The table represent the availability of resemblance with the word "Yes", the non-availability of resemblance with the word "No", the partially resemblance with the word "Yes (partly)". For the unassigned sections the word used is "NA", i.e., not applicable.

Table 6: Availability of Wikipedia category similar to DDC "Science" section

DDC Divisions	DDC Sections	Availability of Wikipedia category
Science 500	Natural sciences & mathematics (500)	Yes
	Philosophy & theory (501)	Yes (Partly)

	Miscellany (502)	No
	Dictionaries & encyclopaedias (503)	No
	Unassigned (504)	NA
	Serial publications (505)	No
	Organizations & management (506)	No
	Education, research & related topics (507)	Yes (Partly)
	Natural history (508)	Yes
	Historical, geographic & persons treatment (509)	No
Mathematics **510**	Mathematics (510)	Yes
	General Principles of Mathematics (511)	No
	Algebra (512)	Yes
	Arithmetic (513)	Yes
	Topology (514)	Yes
	Analysis (515)	Yes
	Geometry (516)	Yes
	Unassigned (517)	**NA**
	Numerical Analysis (518)	Yes
	Probabilities and Applied Mathematics (519)	Yes
Astronomy & Allied sciences **520**	Astronomy (520)	Yes
	Celestial mechanics (521)	Yes
	Techniques, equipment & materials (522)	No
	Specific celestial bodies & phenomena (523)	No
	Unassigned class (524)	NA
	Earth (Astronomical geography) (525)	Yes (Partly)
	Mathematical geography, Cartography (526)	Yes (Partly)
	Celestial navigation (527)	Yes
	Ephemerides (528)	No

	Chronology (529)	Yes
Physics 530	Physics (530)	Yes
	Classical mechanics, solid mechanics (531)	Yes
	Fluid mechanics; liquid mechanics (532)	Yes (Partly)
	Gas mechanics, Pneumatics (533)	Yes (Partly)
	Sound & related vibrations (534)	Yes (Partly)
	Light and related radiation (535)	Yes
	Heat (536)	Yes
	Electricity & electronics (537)	Yes
	Magnetism (538)	Yes
	Modern physics (539)	Yes
Chemistry 540	Chemistry & allied sciences (540)	Yes (Partly)
	Physical Chemistry (541)	Yes
	Techniques, equipment & materials (542)	No
	Analytical Chemistry (543)	Yes
	Unassigned class (544)	NA
	Unassigned class (545)	NA
	Inorganic Chemistry (546)	Yes
	Organic Chemistry (547)	Yes
	Crystallography (548)	Yes
	Mineralogy (549)	Yes
Earth sciences 550	Earth sciences (550)	Yes
	Geology, hydrology & meteorology (551)	Yes
	Petrology (552)	Yes
	Economic geology (553)	Yes
	Earth sciences of Europe (554)	No

	Earth sciences of Asia (555)	No
	Earth sciences of Africa (556)	No
	Earth sciences of North America (557)	No
	Earth sciences of South America (558)	No
	Earth sciences of other areas (559)	No
Palaeontology **560**	Palaeontology (560)	Yes
	Paleobotany; fossil microorganisms (561)	Yes
	Fossil invertebrates (562)	No
	Fossil marine & seashore invertebrates (563)	No
	Fossil mollusks & molluscoids (564)	No
	Fossil arthropods (565)	Yes (Partly)
	Fossil chordates (566)	Yes (Partly)
	Fossil cold-blooded vertebrates; fossil fishes (567)	No
	Fossil Aves (birds) (568)	Yes (Partly)
	Fossil mammals (569)	Yes (Partly)
Biology 570	Biology (570)	Yes
	Physiology & related subjects (571)	Yes
	Biochemistry (572)	Yes
	Specific physiological systems in animals (573)	No
	Unassigned class (574)	NA
	Specific parts of & systems in plants (575)	No
	Genetics & evolution (576)	Yes
	Ecology (577)	Yes
	Natural history of organisms (578)	No
	Microorganisms, fungi & algae (579)	No
Plants 580	Plants (580)	Yes

	Specific topics in natural history (581)	No
	Plants noted for characteristics & flowers (582)	No
	Dicotyledons (583)	No
	Monocotyledons (584)	No
	Gymnosperms; conifers (585)	Yes
	Seedless plants (586)	No
	Vascular seedless plants (587)	No
	Bryophytes (588)	No
	Unassigned class (589)	NA
Animal 590	Animals (590)	Yes
	Specific topics in natural history (591)	No
	Invertebrates (592)	Yes
	Marine & seashore invertebrates (593)	No
	Mollusks & molluscoids (594)	Yes (Partly)
	Arthropods (595)	Yes
	Chordates (596)	Yes
	Cold-blooded vertebrates; fishes (597)	Yes (Partly)
	Birds (598)	Yes
	Mammals (599)	Yes

Out of 100 sections, total number of 93 sections from DDC main class science are assigned, 7 sections are unassigned. After word by word searching in Wikipedia category tree it's found that total 55 categories are found in Wikipedia, out of which 47 categories are completely resemblanced with DDC sections and 8 categories are partly resemblanced. 38 sections are not found in Wikipedia as category.

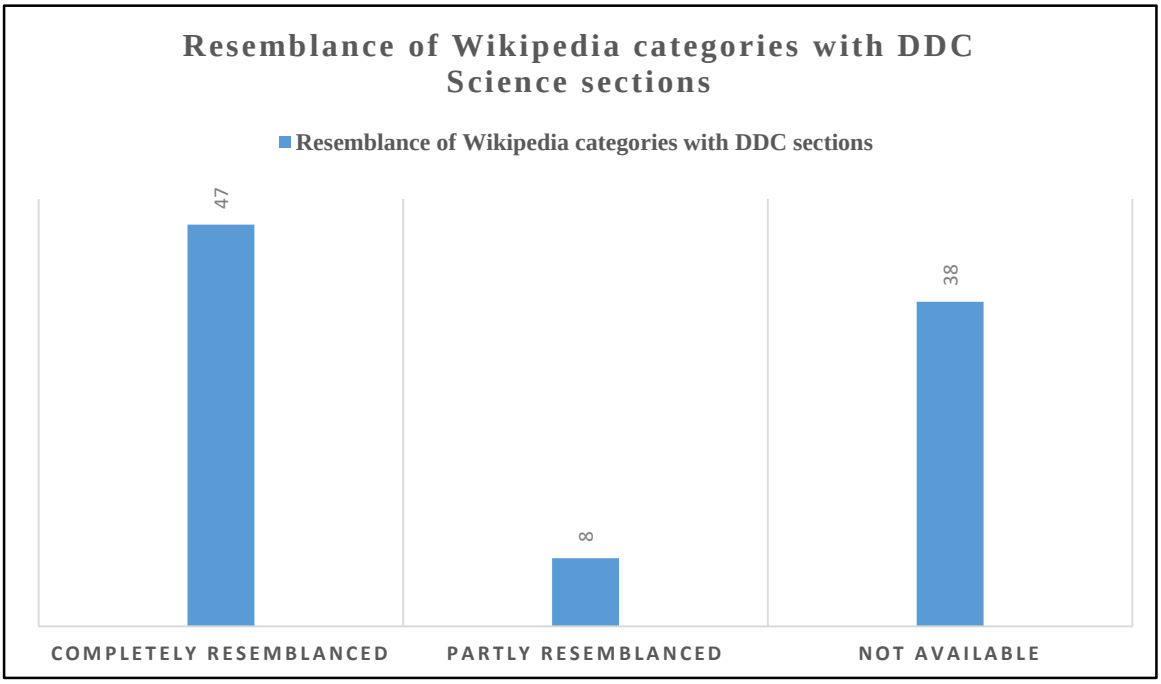


Figure 28: Resemblance of Wikipedia categories with DDC "Science" sections

The above column chart represents the availability of resemblance of Wikipedia category with 90 DDC "science" sections. This overall data division wise is further represented in next few pages.

Division wise data representation for the main class Science

1. Resemblance for the division Science (500)

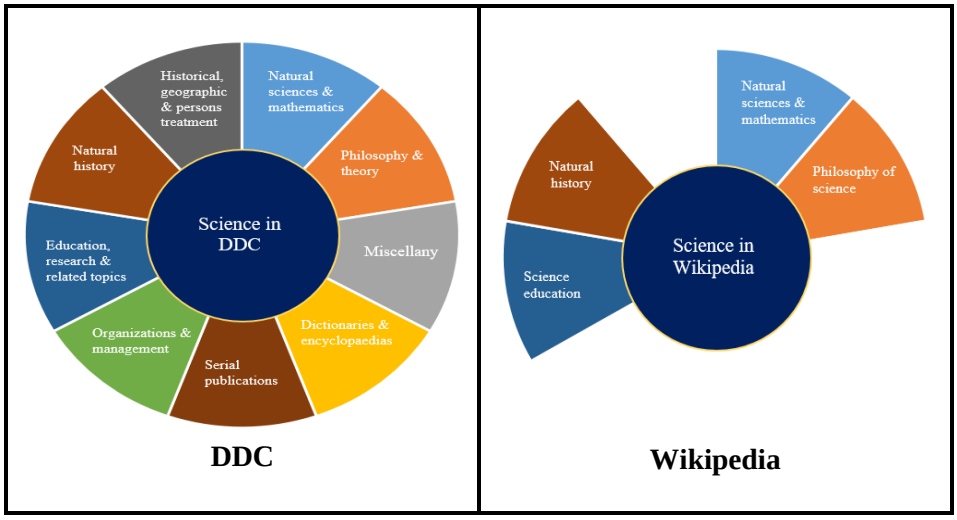


Figure 29: Resemblance for the DDC division Science

2. Resemblance for the division Mathematics (510)

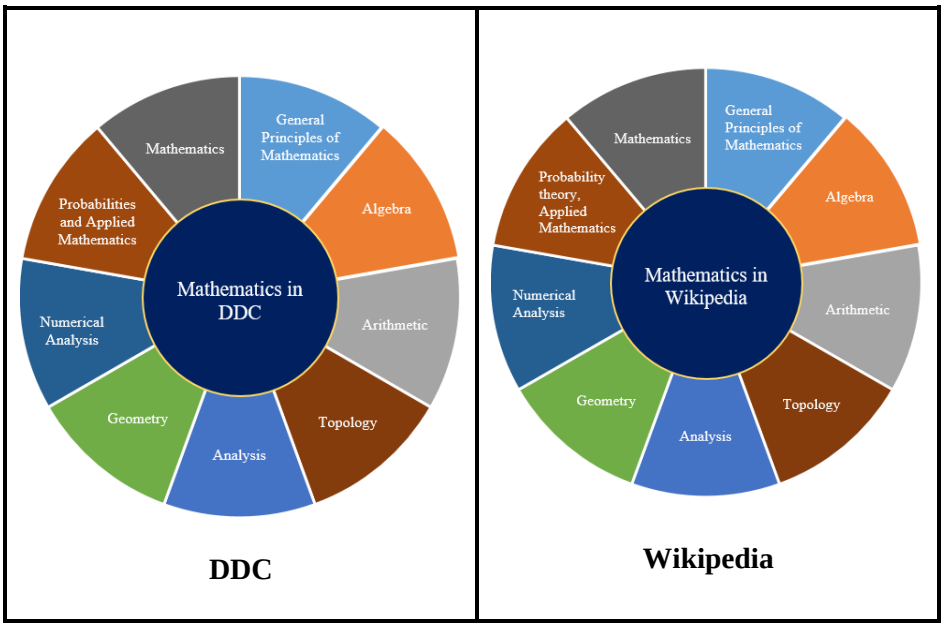


Figure 30: Resemblance for the DDC division Mathematics

3. Resemblance for the division Astronomy (520)

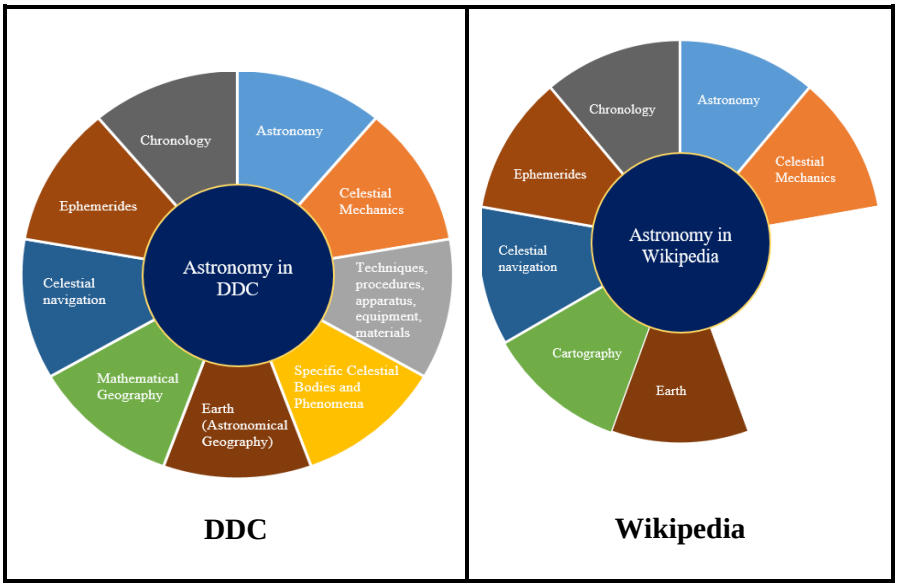


Figure 31: Resemblance for the DDC division Astronomy

4. Resemblance for the division Physics (530)

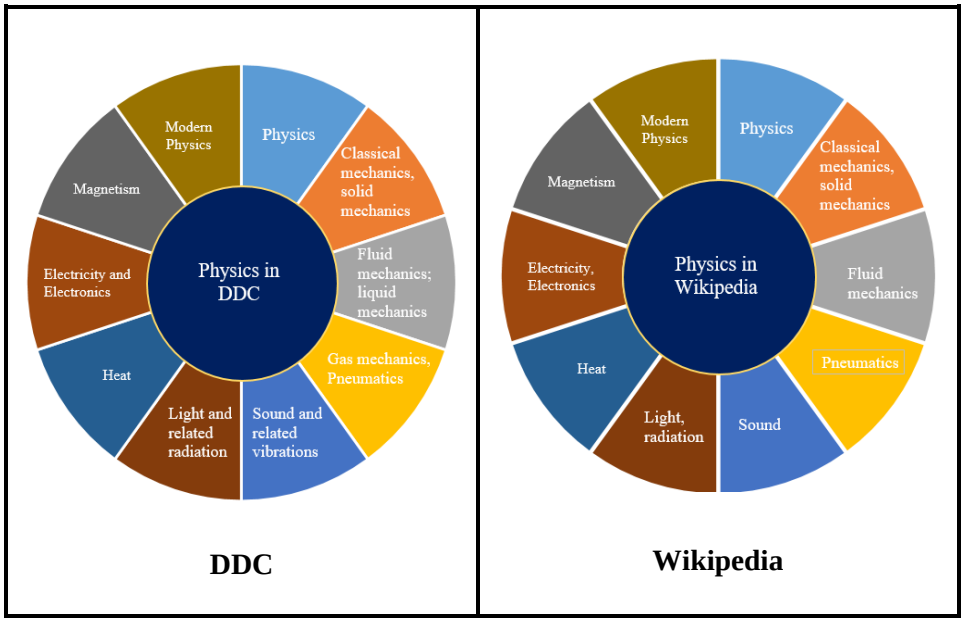


Figure 32: Resemblance for the DDC division Physics

5. Resemblance for the division Chemistry (540)

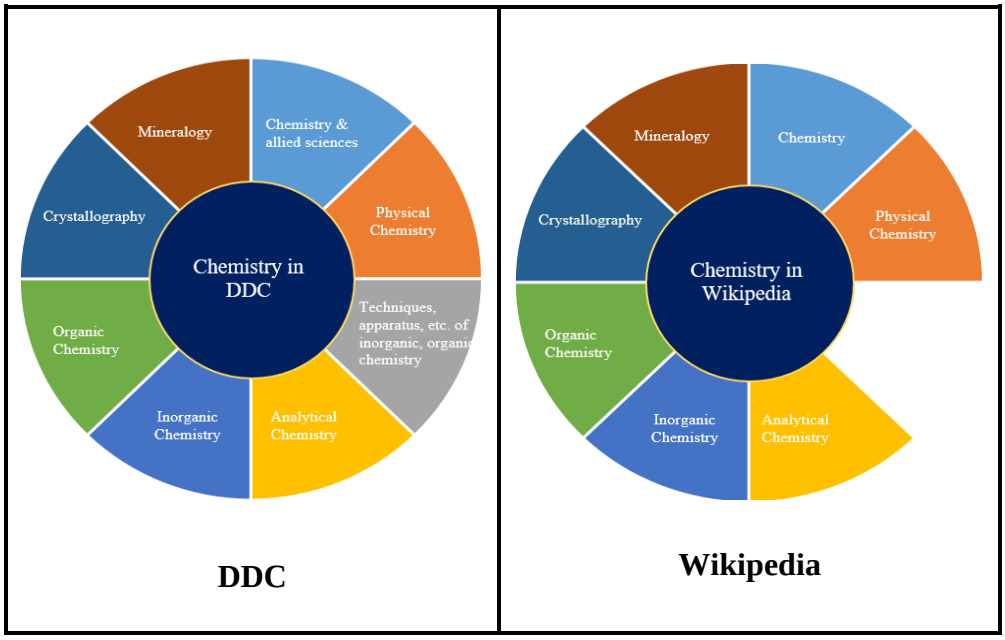


Figure 33: Resemblance for the DDC division Chemistry

6. Resemblance for the division Earth science (550)

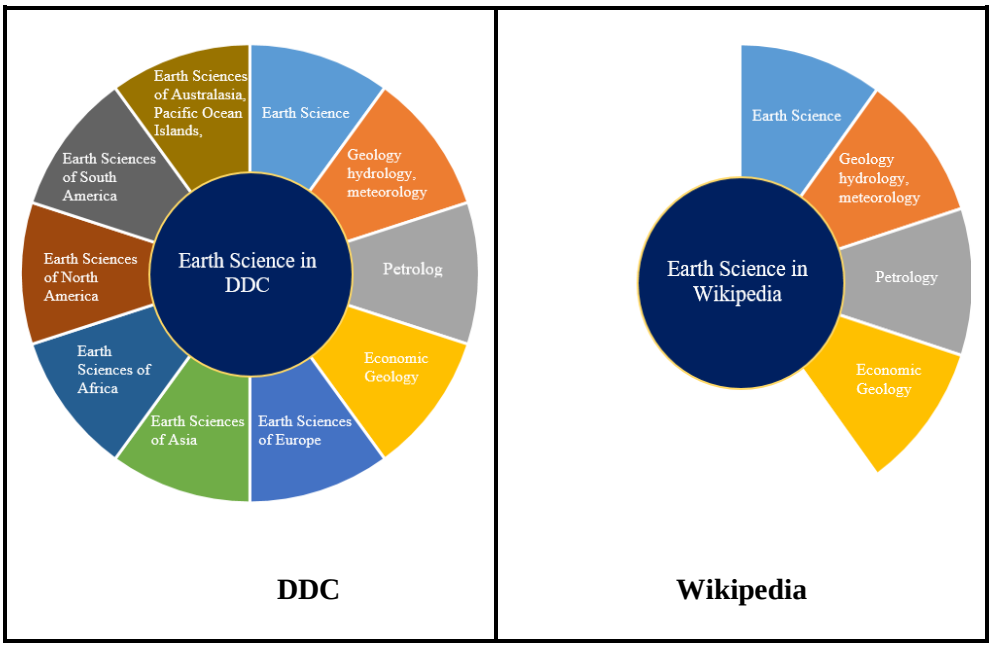


Figure 34: Resemblance for the DDC division Earth Science

7. Resemblance for the division Palaeontology (560)

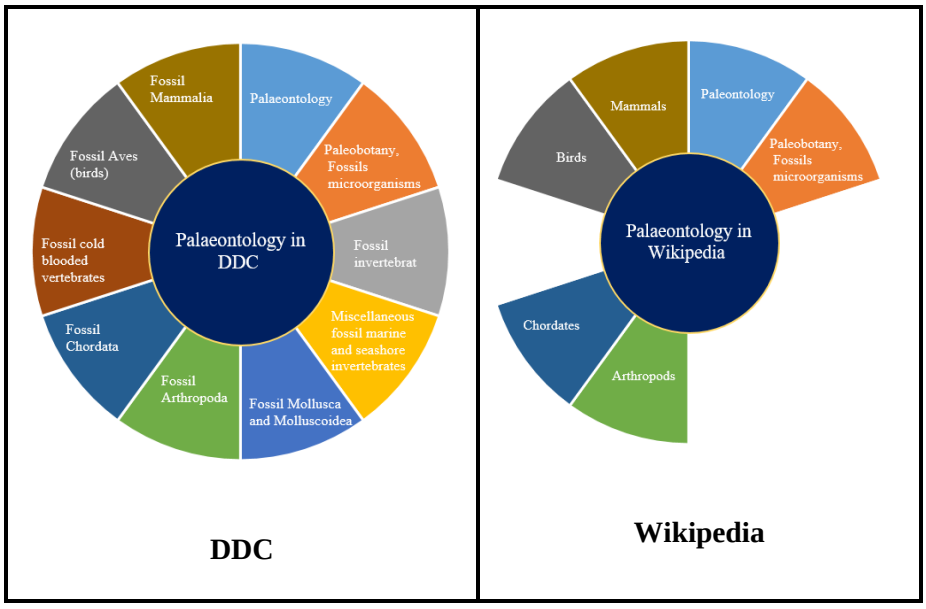


Figure 35: Resemblance for the DDC division Palaeontology

8. Resemblance for the division Biology (570)

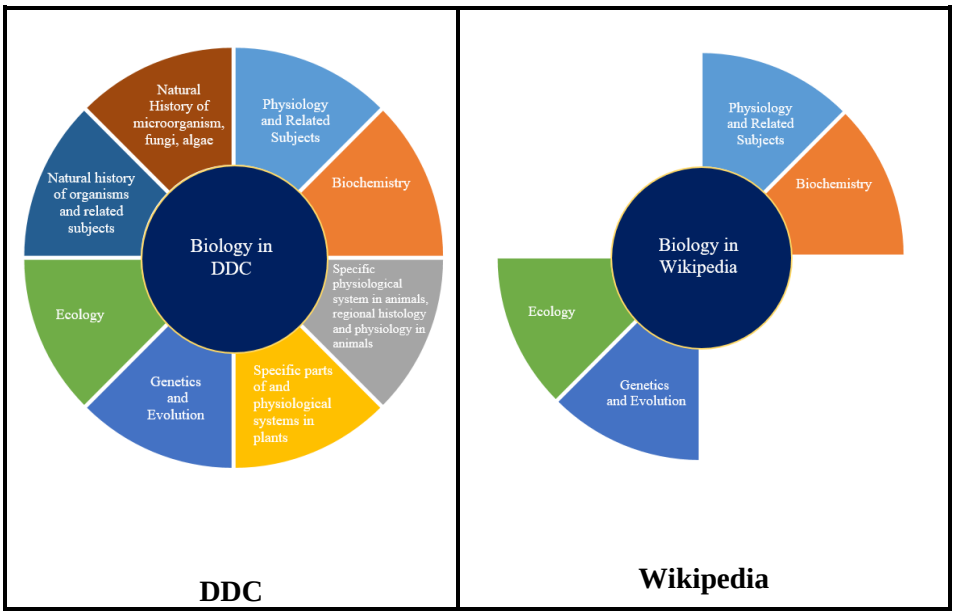


Figure 36: Resemblance for the DDC division Biology

9. Resemblance for the division Plant (580)

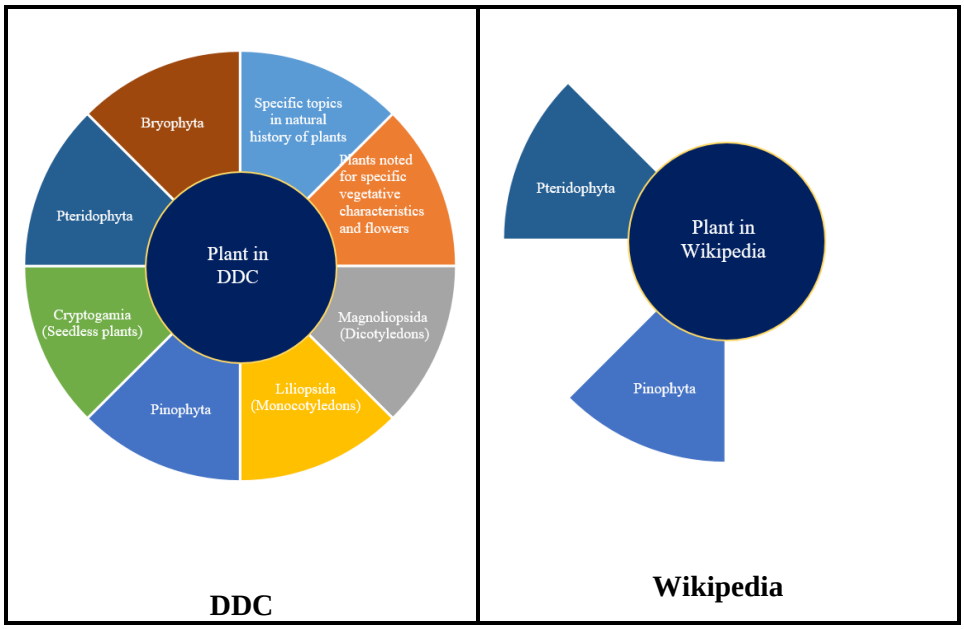


Figure 37: Resemblance for the DDC division Plant

10. Resemblance for the division Plant (590)

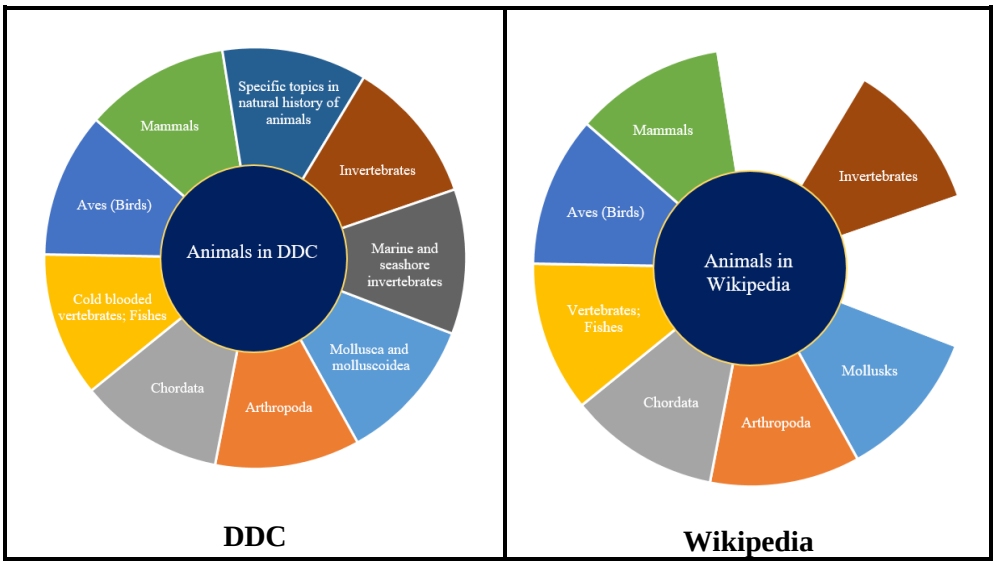


Figure 38: Resemblance for the DDC division Animals

The following column chart represents the status of availability of categories in Wikipedia with the same name of DDC divisions of the main class science. The chart represents the availability or non-availability in whole.

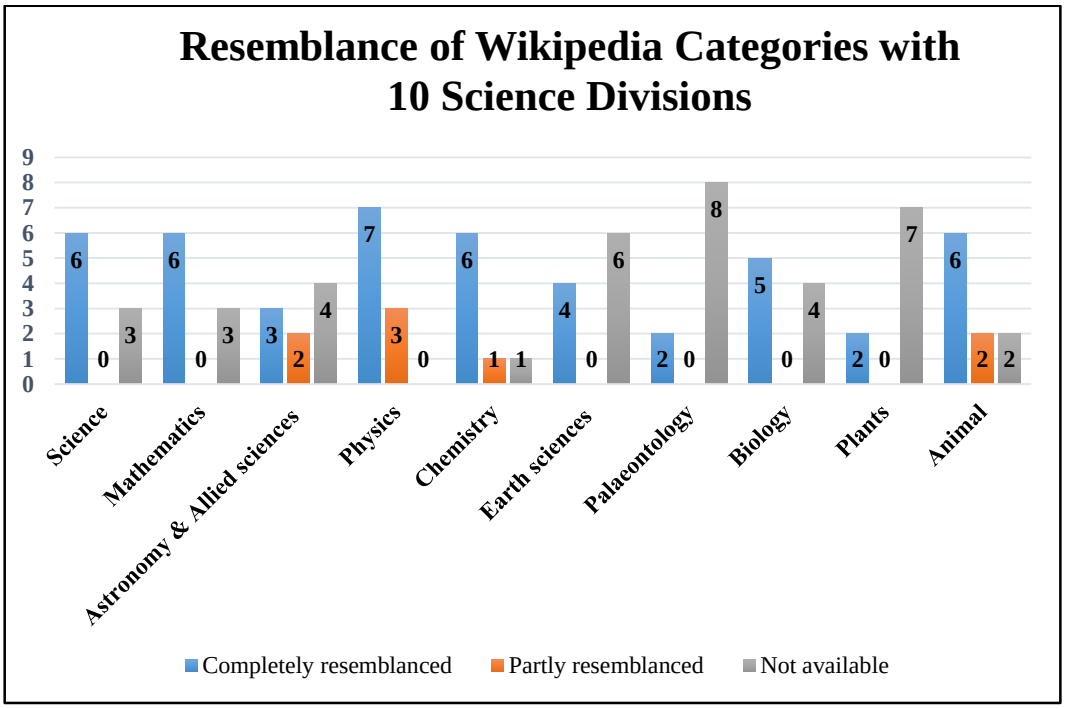


Figure 39: Resemblance of Wikipedia Category with 10 Science divisions

4.2. Data analyses on 'Influence'

Wikipedia classifies its contents, and assigns categories. At the time of writing articles on Wikipedia, the editors have to assign best fitted category to it. In Wikipedia, categories are used to group together pages and/or categories on similar subjects. The Wikipedia categories help readers to find specific category or pages on a specific subject. With the help of Wikipedia category someone can navigate, choose and read articles.

In this part of the study it is tried to identify that whether there is any influence of Dewey Decimal Classification scheme on Wikipedia or not.

Methodology:

1. Collecting the categories similar to the Philosophy divisions, Social Science divisions, and Science divisions. These are the sample categories for this present study.
2. Finding out and noting down the parent categories of these above mentioned sample categories;
3. Marking with 1 mark to the category if it is directly under the parent category similar to Philosophy/Social Science/Science. If the category is one category away from the parent category Philosophy/Social Science/Science, then the category will be given marks 2. Likewise one mark will be increased/decreased depending on the number of categories in between

categories and parent categories. Marking with 0 if the parent category is other than Philosophy/ Social Science/Science. Below is the marking scheme.

Table 7: Marking Scheme for the data analyses on "Influence"

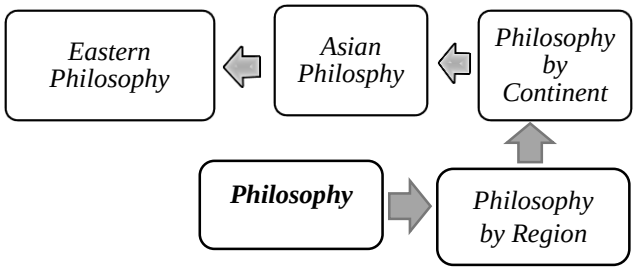
Marking Schemes	**Marks**
When the category is directly under the category Philosophy/Social Science/Science	**1**
When the category is one category away from the parent category similar to Philosophy/Social Science/Science	**2**
When the category is two categories away from the parent category similar to Philosophy/Social Science/Science	**3**
When the category is three categories away from the parent category similar to Philosophy/Social Science/Science	**4**
When Parent category is other than Philosophy/ Social Science/Science	**0**

4.2.1. Parent categories of the category similar to the Philosophy divisions

The below flow charts accommodate all the parent categories of the categories similar to philosophy divisions. Only one category "Philosophical schools and traditions" is directly under the category "Philosophy". In the case of "Metaphysics", "Epistemology", "Ethics" there is one category distance to the parent category "Philosophy". In the case of "Philosophical logic" and "Eastern philosophy" there is three categories distance to the parent category philosophy. The category "Psychology" and "Parapsychology" have parent category "Social Science", not philosophy.

Table 8: Parent categories of the Categories similar to Philosophy divisions

Categories similar to Philosophy divisions.	Parent categories	Points gain
Metaphysics	*Metaphysics* ← *Branches of Philosophy* ← ***Phiosophy***	**2**
Epistemology	*Epistemology* ← *Branches of Philosophy* ← ***Phiosophy***	**2**
Parapsychology	*Parapsychology* ← *Branches of Psychology* ← ***Social Sciences***	**0**
Philosophical schools	*Philosophical Schools and Traditions* ← ***Philosophy***	**1**
Psychology	*Psychology* ← ***Social Sciences***	**0**
Philosophical logic	*Philosophical Logic* ← *Philosophy of Logic* ← *Philosophy by Topic*; ***Philosophy*** → *Branches of Philosophy* ↑ *Philosophy by Topic*	**4**
Ethics	*Ethics* ← *Branches of Philosophy* ← ***Phiosophy***	**2**

Eastern Philosophy		**4**

On the basis of above flow charts points has been given to the categories. Purpose of pointing these categories is to represent the clear picture of parent categories assigned in Wikipedia database. The more the category is away from the parent category "Philosophy", the more it gets points. By following this formula, points has been distributed. The 8 categories (similar to philosophy divisions) get total 15 points. See the below table:

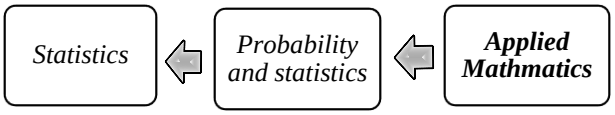
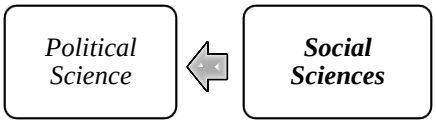
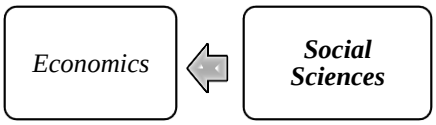
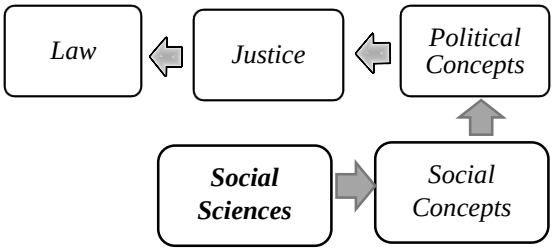
Table 9: Points gained by the Categories similar to Philosophy Divisions

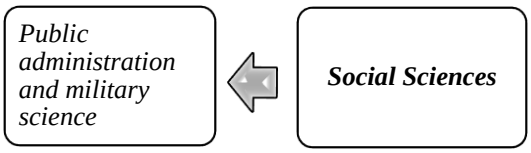
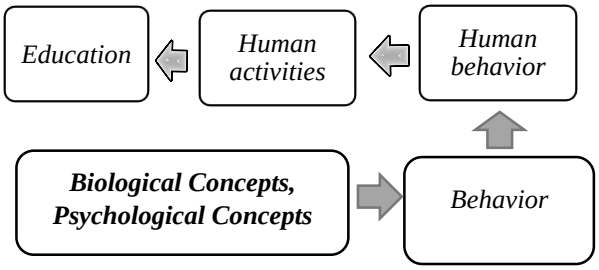
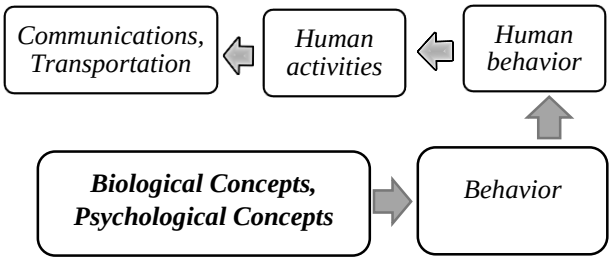
Categories similar to Philosophy Divisions	**Points gained**
Metaphysics	2
Epistemology	2
Parapsychology	0
Philosophical schools	1
Psychology	0
Philosophical logic	4
Ethics	2
Eastern Philosophy	4
Total categories 8 similar to philosophy divisions	Total points15

4.2.2. Parent categories of the category similar to the Social Sciences divisions

The below flow charts accommodate all the parent categories of the categories similar to social science divisions. The three categories, namely “Political Science”, “Economics”, “Public administration and military science” have directly under the category “Social Science”.

Table 10: Parent categories of the Categories similar to Social Sciences divisions

Categories similar to social science divisions.	Parent categories	Points gained
Statistics		0
Political Science		1
Economics		1
Law		4

Public administration and military science		1
Education		0
Commerce, communications, transportation		0
Customs, etiquette, folklore		2

On the basis of above flow charts points has been given to the categories. Purpose of pointing these categories is to represent the clear picture of parent categories assigned in Wikipedia database. The more the category is away from the parent category "Social Sciences", the more it gets points. By following this formula, points has been distributed. The 8 categories (similar to Social Sciences divisions) get total 09 points. See the next table:

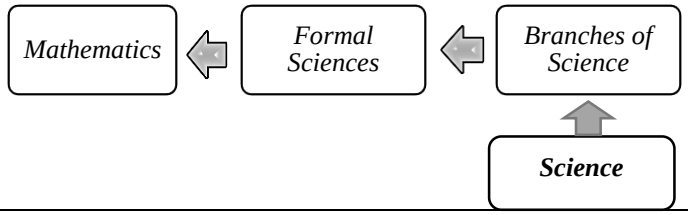
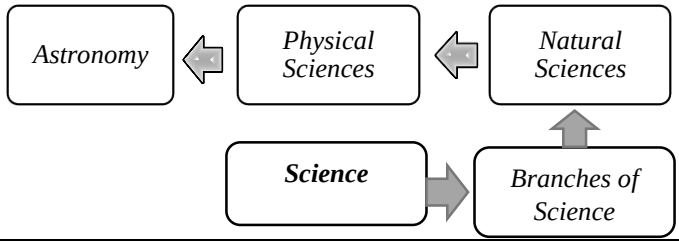
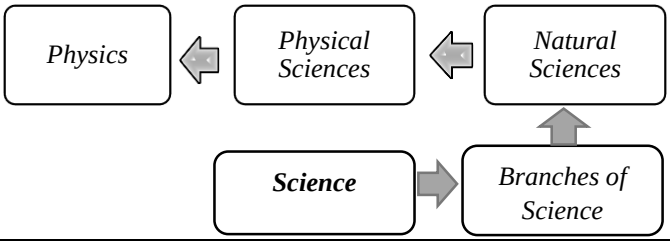
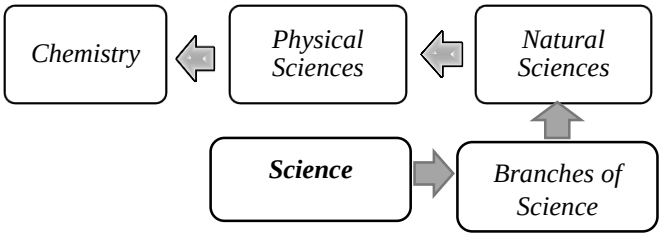
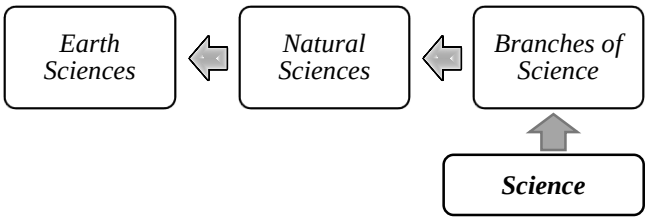
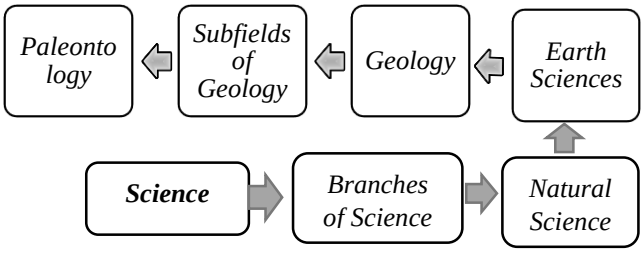
Table 11: Points gained by the Categories similar to Social Sciences Divisions

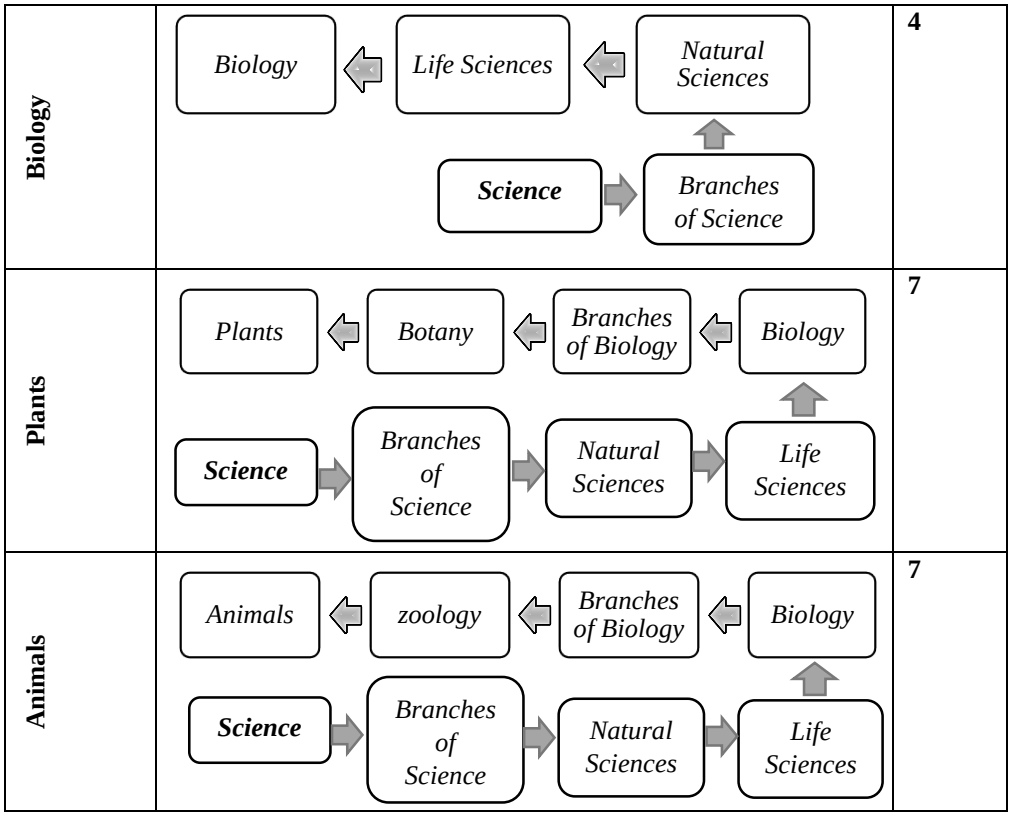
Categories similar to Social Science Divisions	Points gained
Statistics	0
Political Science	1
Economics	1
Law	4
Public administration and military science	1
Education	0
Commerce, communications, transportation	0
Customs, etiquette, folklore	2
Total categories 8 similar to philosophy divisions	Total points 09

4.2.3. Parent categories of the category similar to the Science divisions

This section of the study represents the flow charts of parent categories available in Wikipedia, of the category similar to the science divisions. Here in this encyclopedia, no category is there which is directly under the parent category "Science". The minimum distance between parent category "Science" and the category is 3 and the maximum is 7. The category "Plants" and "Animals" have 7 categories in their hierarchy with the parent category "Science". See the next table which includes all the flow charts.

Table12: Parent categories of the Categories similar to Science divisions

Categories similar to social science divisions.	Parent categories	Points gained
Mathematics		**3**
Astronomy		**4**
Physics		**4**
Chemistry		**4**
Earth sciences		**3**
Paleontology		**6**



On the basis of above flow charts points has been given to the categories. Purpose of pointing these categories is to represent the clear picture of parent categories assigned in Wikipedia database. The more the category is away from the parent category "Science", the more it gets points. By following this formula, points has been distributed. The 8 categories (similar to "Science" divisions) get total 42 points. See the below table:

Table 13: Points gained by the Categories similar to Science Divisions

Categories similar to Science Divisions	**Points gained**
Mathematics	3
Astronomy	4
Physics	4
Chemistry	4

Earth sciences	3
Paleontology	6
Biology	4
Plants	7
Animals	7
Total categories 8 similar to science divisions	**Total points 42**

Findings

Chapter 5: Findings

This chapter of the study is designed to find out the answers of the research questions already written in chapter 3. In this part of the study again the research questions and methodology are rewritten for recalling.

Table 14: Recalling of Research Questions and Methodologies

SN	Research Questions	Methodology
1.	Is there any similarity or resemblance between the categories structure found in Wikipedia with the structure of DDC?	One to one mapping of the subclasses/division under main classes "philosophy", "social sciences", "science" of DDC with Wikipedia
2.	Is the classification of Wikipedia categories into one or more categories influenced by DDC classification system?	Comparing the parent categories of the Wikipedia category with DDC

In the previous chapter data has been analysed by following above methodologies. The study analyses data (sample categories) on the basis of two variables. One is "Resemblance", another one is "Influence". Here the variable "Resemblance" means similarity between DDC and Wikipedia on the basis of availability of the Sections under main classes "philosophy", "social sciences", "science" of DDC in Wikipedia. In Wikipedia categories and articles are classified and are assigned suitable parent categories. In Wikipedia there is a hierarchy of parent, sub and sub-sub categories. Here the variable "Influence" means the impact of the DDC on Wikipedia classification system. Therefore the hierarchy of categories (parent and sub categories) are analysed to find out there is any influence of DDC on Wikipedia available or not.

The next two sections of the chapter, will find out, whether there is any resemblance or similarity between DDC and Wikipedia or not and there is any influence of DDC on Wikipedia or not.

5.1. Resemblance between two knowledge organization tools DDC and Wikipedia

To analyse the resemblance or similarity, the sections under the main classes "Philosophy", "Social Sciences", "Science" mentioned in DDC 23rd edition were noted down. The following table represents the assigned sections of DDC and the availability of categories of same name in Wikipedia.

Table 15: Resemblance between two knowledge organization tools DDC and Wikipedia

	Sections assigned in DDC	**Fully similar categories in Wikipedia**	**Partially Similar categories in Wikipedia**	**Dissimilar categories in Wikipedia**
Philosophy	89	58 (65%)	11 (12%)	20 (23%)
Social Sciences	90	46 (51%)	13 (15%)	31 (34%)
Science	93	47 (50%)	08 (9%)	38 (41%)
	272	151 (55%)	32 (12%)	89 (33%)

The above data and the below pie charts show that Wikipedia contains maximum similar categories for the subject "Philosophy", and minimum similar categories for the subject "Science". Maximum dissimilarity of categories are found for the Science subject. Minimum dissimilarity of categories are found for the subject Philosophy.

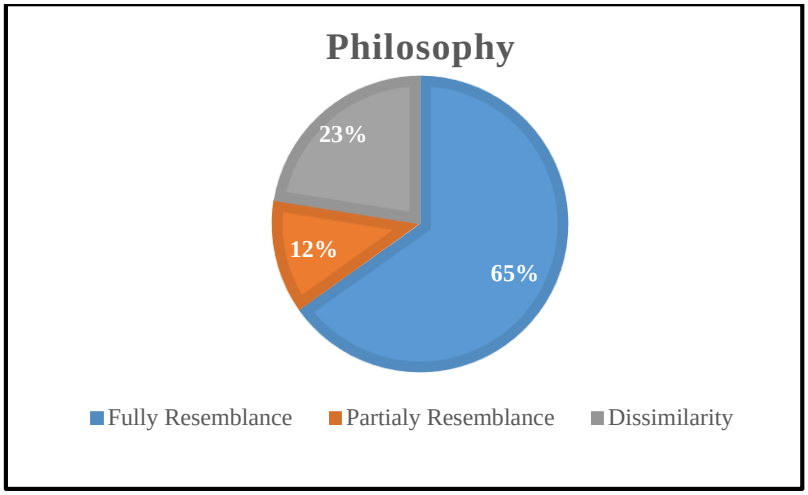


Figure 40: Resemblance in DDC main class Philosophy

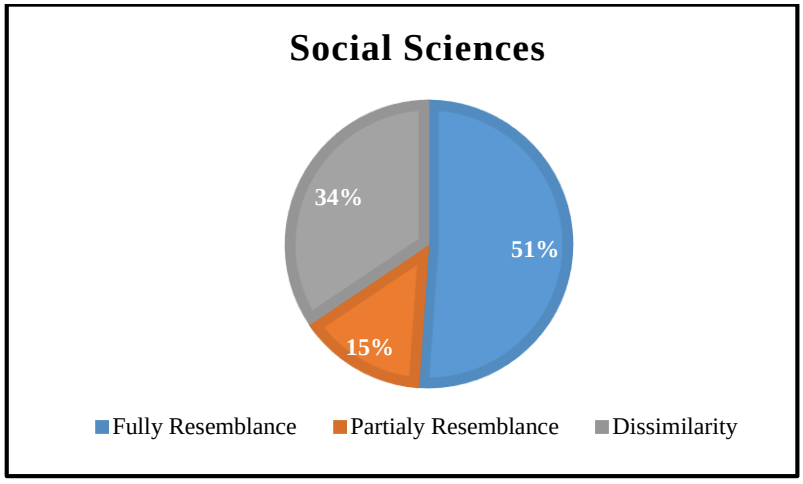


Figure 41: Resemblance in DDC main class Social Sciences

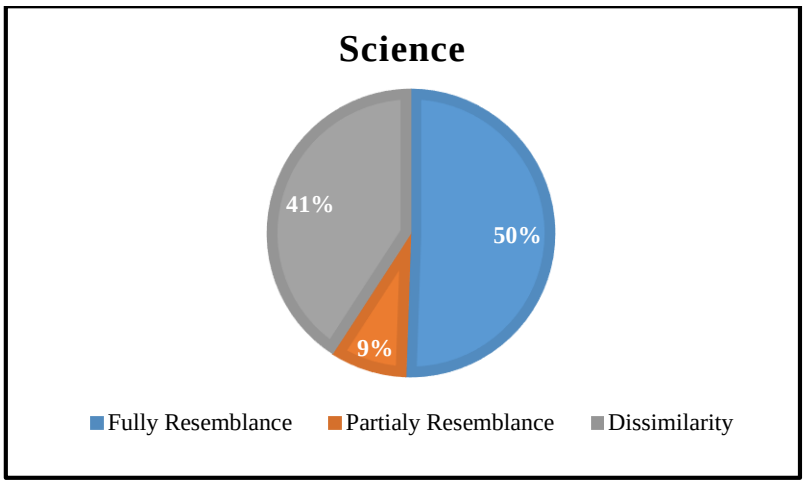


Figure 42: Resemblance in DDC main class Science

The present study analysed Wikipedia Data for three subjects philosophy, social sciences, and science. Total 272 sections are noted from DDC 23rd edition from the three main subject "Philosophy", "Social Sciences", "Science". Out of 272 sections, **151 sections (55%) were found in Wikipedia as category. 89 sections (33%) were not** found in Wikipedia. There is partial similarity of categories for 12%, it means 32 categories were found partially similar to the collected sections. See the below pie chart:

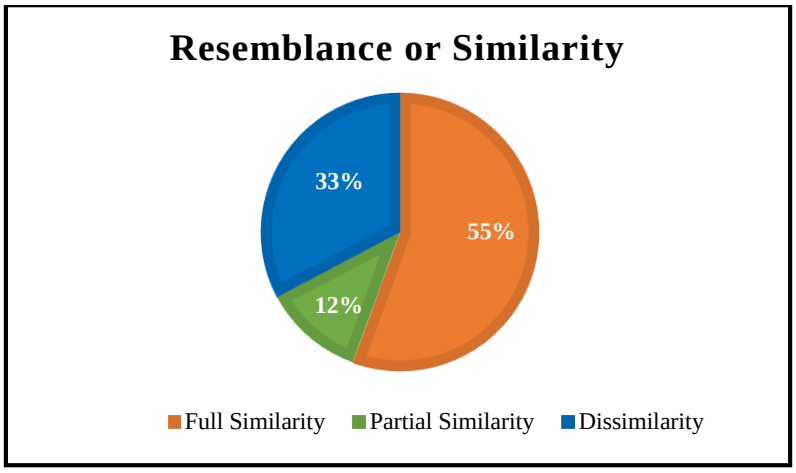


Figure 43: Resemblance between Wikipedia categories and DDC

Findings on the basis of the variable "Resemblance"

1) **55% Similar Categories:** For three main classes namely "Philosophy", "Social Sciences", "Science", Wikipedia contains 55% **similar categories.** Out of selected 272 sections of DDC, 151 sections were found in Wikipedia as category.
2) **33%Dissimillar Categories:** Out of selected 272 sections of DDC, 89 sections were not found in Wikipedia as category. The percentage of dissimilarity is 33%.

3) **12% Partially Similar Categories:** Few categories were there which are not fully but partially similar with the DDC sections.
4) **Maximum and Minimum Similar categories:** Wikipedia contains maximum similar categories for the subject “Philosophy” which is 65%, and minimum similar categories for the subject “Science” which is 50%. The encyclopedia contains 51% similar categories for the subject “Social Sciences”.

5.2 Influence of DDC Classification Scheme on Wikipedia

It was decided to find out whether there is any influence of DDC on Wikipedia or not by comparing the parent categories of the Wikipedia category with DDC. For recalling, the methodology is again write down here:

Methodology:

- Collecting the categories similar to the Philosophy divisions, Social Sciences divisions, and Science divisions. These are the sample categories for this present study.
- Finding out and noting down the parent categories of the sample categories;
- Marking with 1 mark to the category if it is directly under the parent category similar to Philosophy/Social Sciences/Science. If the category is one category away from the parent category Philosophy/Social Sciences/Science, then the category will be given marks 2. Likewise one mark will be increased/decreased depending on the number of categories in between categories and parent categories. Marking with 0 if the parent category is other than Philosophy/ Social Sciences/Science. Below was the marking scheme.

Table 16: Recalling the Marking Scheme for the data analyses on "Influence"

Marking Scheme	**Points**
When the category is directly under the category Philosophy/Social Sciences/Science	1
When the category is one category away from the parent category similar to Philosophy/Social Sciences/Science	2
When the category is two categories away from the parent category similar to Philosophy/Social Sciences/Science	3
When the category is three categories away from the parent category similar to Philosophy/Social Sciences/Science	4
When the Parent category is other than Philosophy/ Social Sciences/Science	0

On the basis of above marking scheme, points was given to the sample categories similar to Philosophy/Social Sciences/ Science divisions. It was to find out is there any influence of DDC classification scheme on Wikipedia category structure or not. The above marking scheme says if there is any influence then points would be "1" or near to 1. If the point is "0" then it means there is no influence of DDC scheme on Wikipedia category structure. The result is as follows:

Table 17: Total points gained by Philosophy

Categories similar to Philosophy Divisions	**Points gained**
Metaphysics	2
Epistemology	2
Parapsychology	0
Philosophical schools	1
Psychology	0
Philosophical logic	4
Ethics	2
Eastern Philosophy	4
Total points	**15**

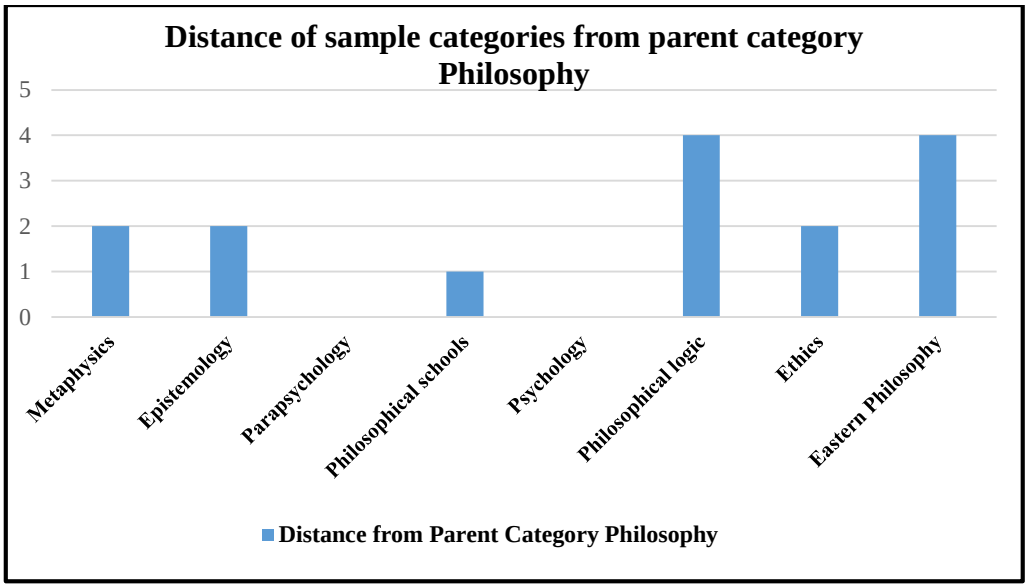


Figure 44: Distance of sample categories from parent category Philosophy

For the categories similar to philosophy divisions the result is given in above table. Only the category "Philosophical schools" get 1 point. The maximum point is 4 gained by "Philosophical logic" and "Eastern philosophy". There are also point 0 gained by "Parapsychology" and "Psychology". It means philosophy is not parent category of these two categories. The total point is 15.

Table 18: Total points gained by Social Sciences

Categories similar to Social Science Divisions	**Points gained**
Statistics	0
Political Science	1
Economics	1
Law	4
Public administration and military science	1

Education	0
Commerce, communications, transportation	0
Customs, etiquette, folklore	2
	Total points 09

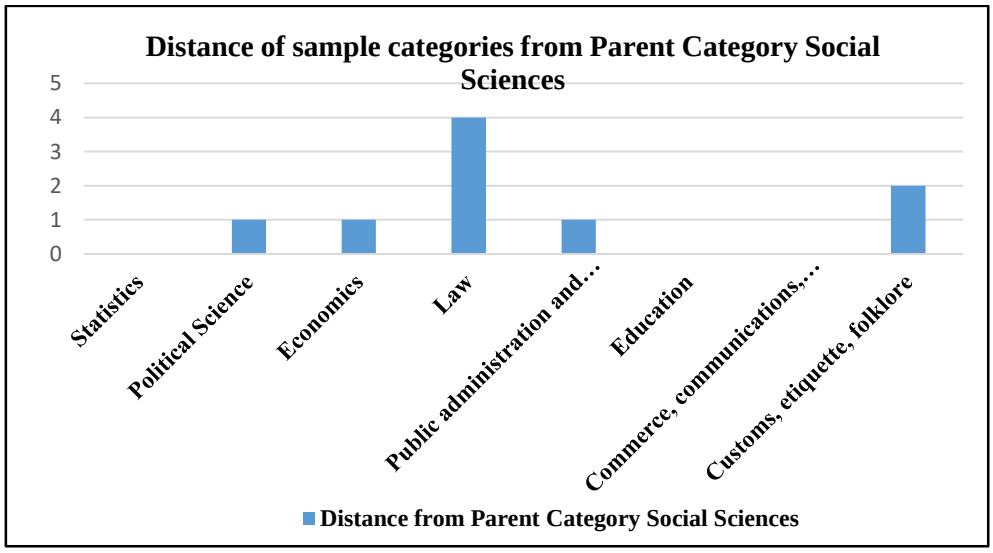


Figure 45: Distance of sample categories from parent category Social Sciences

For the categories similar to social sciences divisions the result is given in above table. There are three zero's. The categories are "Statistics", "Education", "Commerce, communications, and transportation". It means not all the sample categories similar to Social Science Divisions have social sciences as parent category. The point "1" is gained by three categories, "Political science", "Economics", "Public administration and military science". It means these categories are directly under the parent category "Social sciences". The highest category is "4", gained by "Law". The total point is 09.

Table 19: Total points gained by Science

Categories similar to Science Divisions	Points gained
Mathematics	3
Astronomy	4
Physics	4
Chemistry	4
Earth sciences	3
Paleontology	6
Biology	4
Plants	7
Animals	7
	Total points 42

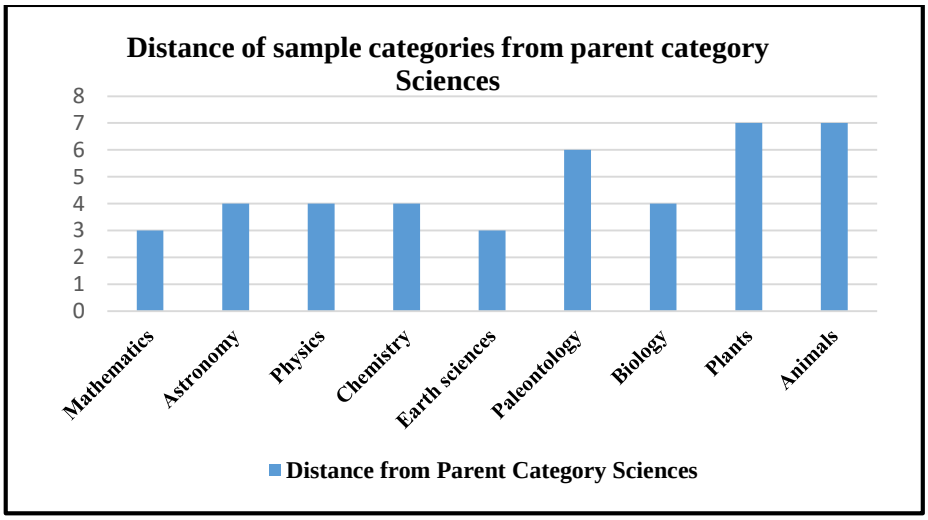


Figure 46: Distance of sample categories from parent category Science

For the categories similar to science divisions the result is given in above table. The result says there is no point "1". It means none of the sample categories is under science category directly. There is no point "2". The minimum point is "3", gained by "Mathematics", "Earth Sciences". The maximum point is "7" gained by "Plants", "Animals". There is no point "0". It means all the sample categories are under parent category "Science". But here the total point is 42.

Findings on the basis of the variable "Influence"

1) To find out the result a marking scheme was designed, which indicates that if the category hierarchy of a particular subject is influenced by DDC classification scheme then it must have smallest point, with no zero.
2) Two zeroes were found in the sample categories collected similar to DDC philosophy divisions. The total point is 15.
3) Three zeroes were found in the sample categories collected similar to DDC social sciences divisions. Here the total point is 09.
4) Apparently the category structure of the subject "Science" is found not much influenced by DDC classification scheme, as the total point here is 42. But all the sample categories collected similar to DDC sciences divisions are having parent category "Science". There is no zero. The lowest point is "3" gained by "Mathematics" and "Earth Science" and the highest point is "7", gained by "Plants" and "Animals".
5) Out of 25 sample Wikipedia categories (similar to DDC Philosophy, Social Sciences, Science divisions) 5 categories are not having the same parent categories as described in DDC. Those 5 categories (with 0 points) are Parapsychology, Psychology, Statistics, education, Commerce, communications, transportation. Rest of the 20 categories are having similar parent categories as described in DDC (see next pie chart).

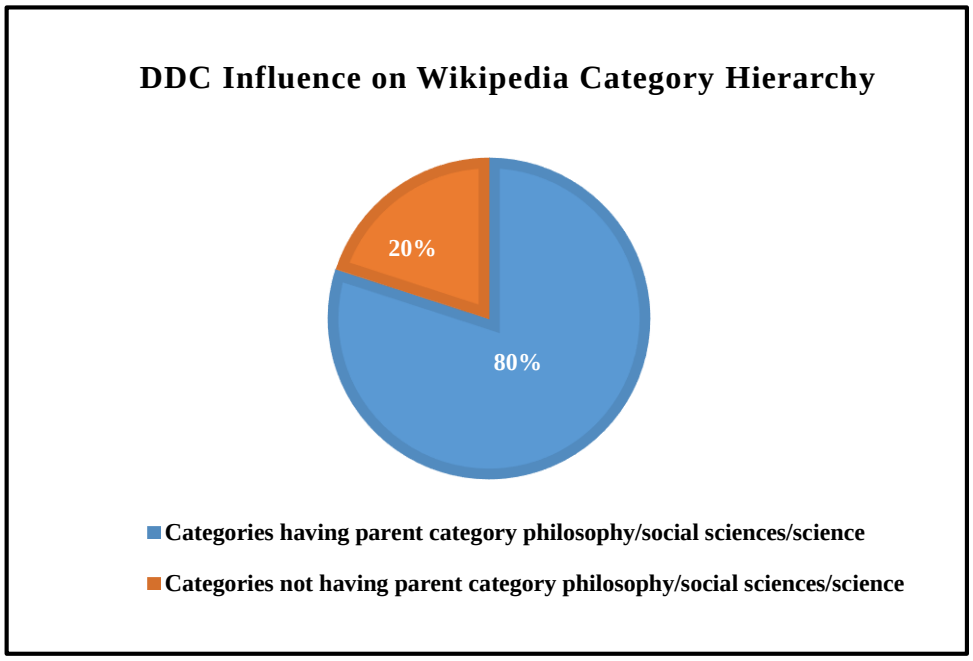


Figure 47: DDC Influence on Wikipedia Category hierarchy

5.3. Findings of Research Questions

The following table accumulate the findings of the research questions designed for the study.

Table 20: Research questions and findings

SN	Research Questions	Methodology	Findings
1.	Is there any similarity or resemblance between the categories structure found in Wikipedia with the structure of DDC?	One to one mapping of the subclasses/division under main classes "philosophy", "social sciences", "science" of DDC with Wikipedia	Yes, there is a similarity between Wikipedia category structure and DDC class structure. Wikipedia contains **55% similar categories.** Out of selected **272** sections of DDC, **151** sections

			were found in Wikipedia as category.
2.	Is the classification of Wikipedia categories into one or more categories influenced by DDC classification system?	Comparing the parent categories of the Wikipedia category with DDC	Wikipedia is somehow influenced by DDC classification system. Out of 25 DDC divisions, 5 divisions are not found in Wikipedia with the same parent category as DDC. **80 % of total sample categories are influenced by DDC class divisions**, as they are having similar parent categories described as in DDC.

Conclusions

Chapter 6: Conclusions

To organize Wikipedia content, to give its users feasible searching facilities among the articles and categories, the editors are given a guideline to classify its articles and categories by assigning appropriate category names. This process may be named as **"Social Classification", "Collaborative Tagging", "Social Indexing" or "Folksonomy**", where all the articles are organised by wide range of topics from politics to science, news events to literature. The continuous growth of its content make this encyclopedia as the reflection of human knowledge. In recent times Wikipedia category structure is being used for document classification, information retrieval, to improve text classification. Many researchers are working on Wikipedia's thesaurus like features. The class structure in DDC and the category structure in Wikipedia are maintained differently on the basis of different scheme of classification. DDC is maintained by classification experts and Wikipedia class structures are created by Wikipedia editors/ Wikipedians among the general people. But both are using for classification. Therefore it was to find out is there any influence and similarity of any classification scheme in this category system or not.

The present study was started with the aim to find out Resemblance and influence of DDC Classification scheme on Wikipedia category structure. Because of English Wikipedia's huge collection of articles and categories, three subjects were chosen as sample of the study. These three subjects are Philosophy, Social Sciences, and Science. To find "**resemblance**" between Wikipedia category structure and DDC class structure one to one mapping was conducted. To find **"influence"** of DDC on Wikipedia Category Structure, the parent categories of sample categories were noted down.

The analyses on the basis of "Resemblance" variable give a result where the sample categories of three subjects are resemblanced with DDC class names. The percentage of resemblance is 55. Out of selected 272 sections of DDC, 151 sections were found in Wikipedia as category.

The analyses on the basis of variable "Influence" shows that the sample categories are having parent categories similar to DDC class structure. The hierarchy of Wikipedia categories sub-categories is influenced by DDC. Out of 25 sample Wikipedia categories (similar to DDC Philosophy, Social Sciences, Science divisions) 5 categories are not having the same parent categories as described in DDC. Here the influence of DDC class structure on Wikipedia category structure is 80%.

The present study founds Wikipedia's similarity with DDC classification scheme and in many places the influence of DDC class structure is found in Wikipedia. Therefore it may be concluded that somehow accidently or incidentally Wikipedia category structure is similar with DDC class structure and there is an influence of DDC in Wikipedia. Further research is required to know the influence of other scheme on this mass generated encyclopedia.

6.1. Scope for Further Research

During the research works, the present study finds following thesaurus like elements in Wikipedia:

Table 21: Thesaurus like features in Wikipedia

Elements	**Thesaurus like features**
Redirects	"see" references by linking synonyms to preferred terms.

Disambiguation pages	Homonyms.
The poly hierarchical category structure	Broader and narrower term relationships
Links between pages	Related term indicators

Because of these thesaurus like features many researchers uses Wikipedia for metadata enrichment, text clustering, and classification. When this encyclopedia is being used to classify documents, resources, then it is necessary to search that is there any influence of any classification scheme on Wikipedia or not. The present study found Wikipedia's similarity with DDC classification scheme. But the research work may be done for other classification scheme also.

6.2. Recommendation

Wikipedia is a free online encyclopedia that anyone can edit. The statistics says that in Wikipedia, there are 2 edits in every second, average 537 new articles are written per day. As on date only English Encyclopedia includes 6,820,061 articles and around 20 lacks total categories, as on September, 2024. Therefore it is a huge task to classify and organise the giant collection of this encyclopedia. Although encyclopaedia's articles are classified by categorising where vast majority of pages belong to at least one category. There are few recommendations:

- The entire collection of the encyclopedia may be classified according to any popular classification scheme.
- There is many loop connections between all the connected categories. The collection of categories may be organised in tree like hierarchy.

- As this encyclopedia is the biggest encyclopedia of the world, anyone can lost in its collection. Therefore it is recommended to design proper site map of categories (without hidden loops) to guide its users all over the world.

References

References

Afterword, S. (2007, March 21). *The Wikipedia Factor in US intelligence.* https://fas.org/publication/the_wikipedia_factor_in_us_int/

Arias, M. L. (2007). Internet Law- Wikipedia: the free online encyclopedia and its use as court source. In *Internet Business Law Services*. https://www.druganddevicelawblog.com/2017/01/pitfalls-of-judges-lawyers-and-experts-citing-wikipedia.html

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. G. (2007). DBpedia: A Nucleus for a Web of Open Data. *Lecture Notes in Computer Science, 6,* 722-735. https://doi.org/10.1007/978-3-540-76298-0_52.

Bhandari, P. (2023). *What Is Qualitative Research? Methods & Examples.* Scribbr. https://www.scribbr.com/methodology/qualitative-research/

Brown, A. R. (2011). Wikipedia as a data source for political scientists: accuracy and completeness of coverage. *Political Science and Politics, 44*(2), 339-343. http://adambrown.info/docs/research/brown-2011-wikipedia-as-a-data-source.pdf

Butler, D. (2008). Publish in Wikipedia or perish. *Nature*. https://doi.org/10.1038/news.2008.1312

Cail, P., Feng, Y., Jial, Y., Wang, Y., Jin, X., & Cheng, X.(2017). *Coarse to fine: diffusing categories in Wikipedia* [Conference session]. 26th International Conference on World Wide Web. https://doi.org/10.1145/3041021.3054235

Category: Main topic classifications. (2019, Feb. 2024). In *Wikipedia.* https://en.m.wikipedia.org/wiki/Category:Main_topic_classifications.

Chahine, C. A., Chaignaud, N., Kotowicz, P., & Pécuchet, J. (2011). Conceptual indexing of documents using Wikipedia. *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (pp. 195-202). IEEE Xplore Digital Library. https://ieeexplore.ieee.org/document/6040518

Chen, Z., Cao, J., Song, Y., Zhang, Y., & Li, J. (2010). Web video categorization based on Wikipedia categories and content-duplicated open resources. *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference* (pp.1107-1110). ACM Digital Library. https://doi.org/10.1145/1873951.1874162

Civil Marriage Act. (2017). In *Parliament of Canada*. https://www.parl.ca/LegicInfo/BillDetails.aspx?billId=1585203&View=10

Clauson, K. A., Polen, H. H., Boulos, M. N. K., & Dzenowagis, J. H. (2008). Scope, completeness, and accuracy of drug information in Wikipedia. *Ann*

Pharmacother, 42, 1814-1821. http://ils.unc.edu/bmh/tmp/706/Clauson2008-WikipediaDrugInfo_review.pdf.

Colgrove, C., Neidert, J., & Chakoumakos, R. (2011, December 11). Using network structure to learn category classification in Wikipedia. http://snap.stanford.edu/class/cs224w2011/proj/colgrove_Finalwriteup_v1.pdf.

Crovitz, D., & Smoot, W. S. (2009). Wikipedia: friend, not foe. *English Journal, 98*(3), 91–97. https://www.nytimes.com/learning/teachers/archival/EnglishJournalArticle2.pdf

Dewey summaries. (2024). In *OCLC*. Retrieved September 4, 2024, from https://www.oclc.org/content/dam/oclc/dewey/resources/summaries/deweysummaries

Eyal, B. (2018). Disciplines and the categorization of scientific truth: the case of social sciences in the Hebrew Wikipedia. *Journal of Sociocybernetics, 42,* 20-37.https://www.researchgate.net/publication/329949670_Disciplines_and_the_Categorization_of_Scientific_Truth_The_Case_of_Social_Sciences_in_the_Hebrew_Wikipedia.

Ford, H. & Sen, S. (2013). Getting to the source: where does Wikipedia get its information from?. *Information & Culture, 47*(2), 233-254. from http://www.opensym.org/wsos2013/proceedings/p0203-ford.pdf.

Ghosh, P. & Mondal, T. K. (2018). How to write articles in Wikipedia: a tutorial. *Librarian, 23(*1), 50-62.

Ghosh, P. & Mondal, T. K. (2018). Library Wikipedia collaboration becomes the new paradigm library services. In Dr. Susmita Chakraborty (Ed.), *Transforming the society: libraries as the torch-bearer of change* (pp.220-227). Paper presented at World Library & Information Congress 2018 Satellite Session (Pre-Conference): IFLA.CU 2018. Kolkata: Department of Library & Information Science, University of Calcutta.

Ghosh, P. & Mondal, T. K. (2018). Wikipedia as a digital reference source in public libraries of India. In Prof. S. K. Das, Prof. S. K. Chatterjee, Prof. G. Maity & Dr. T. K. Mondal (Eds.), *International seminar on rejuvenating public library services through digital reference sources* (pp. 415-418). Kolkata: Department of Library & Information Science, Jadavpur University.

Ghosh, P. & Mondal, T. K. (2018, Nov. 27-29). *Wikipedia as an educational aid for library and information science subject: an analysis* [Conference session]. 28th IASLIC National Seminar, Indian Association of Special Libraries & Information Centres, Kolkata, Visva-Bharati, Shantiniketan.

Giles, J. (2005). Internet encyclopedias go head to head. *Nature,* 438. http://inspercom.org/wpcontent/uploads/2015/06/GILES_Internet-encyclopaedias-go-head-tohead2005Cit.496_Junho-de-2015.pdf

Halavais, A., & Lackaff, D. (2008). An analysis of topical coverage of Wikipedia. *Journal of Computer-Mediated Communication, 13,* 429–440. http://onlinelibrary.wiley.com/doi/10.1111/j.10836101.2008.00403.x/full

Hasty, R. T., Garbalosa, R. C., Barbato, V. A., Valdes JR, P. J., Powers, D. W., Hernandez, E., John, J. S., Suciu, G., Qureshi, F., Popa-radu, M., Jose, S. S., Drexler, N., Patankar, R., Paz, J. R., King, C. W., Gerber, H. N., Valladares, M. G., & Somji, A. (2013). Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions. *J Am Osteopath Assoc, 114(*5), 368-73. https://www.cecity.com/aoa/jaoa_mag/2014/may_14/368.pdf.

Higashinaka, R., Sadamitsu, K., Saito, K., Makino, T., & Matsuo, Y. (2012). *Creating an extended named entity dictionary from Wikipedia*. 24th International Conference on Computational Linguistics *(*pp.1163-1178). https://www.researchgate.net/publication/270877528_Creating_an_Extended_Named_Entity_Dictionary_from_Wikipedia

Hoffart, J., Suchanek, F. M., Berberich, K., & Weikum, G. (2013). Yago2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence, 194,* 28-61. http://www.sciencedirect.com/science/article/pii/S0004370212000719

Holloway, T., Bozicevic, M., & Borner, K. (2007). Analyzing and visualizing the semantic coverage of Wikipedia and its authors. *Wiley Periodicals, 12*(3). http://onlinelibrary.wiley.com/doi/10.1002/cplx.20164/pdf

Kaptein, R., & Kamps, J. (2013). Exploiting the category structure of Wikipedia for entity ranking. *Artificial intelligence, 194,* 111-129. https://reader.elsevier.com/reader/sd/pii/S0004370212000732?token=B912507B0E8B6A2A712A2C2CEB3F2CE4F22361C3400CDE87707B0CCC4C7A836E4D315EF8BC244B6962EC5CD52D5A613C

Kumer, K. (2010). *Theory of Classification* (Second Edition). New Delhi: Vikas Publishing House Pvt Ltd.

Legally India. (2010, October 23). 'Offensive' SC judgment cites Wikipedia to define legal term. https://www.legallyindia.com/the-bench-and-the-bar/offensive-sc-judgment-cites-wikipedia-to-define-legal-term-20101023-1445#:~:text=gavel%20A%20Supreme%20Court%20of,stand%E2%80%9D%20and%20%E2%80%9Ckeep%E2%80%9D.

Lehman, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. M., Hellmann, S., Morsey, M., Kleef, P. V., Auer, S. & Bizer, C. (2012). DBpedia a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web, 1*. http://www.semantic-web-journal.net/system/files/swj499.pdf

List of most visited website. (2024, October 29). In *Wikipedia*. https://en.wikipedia.org/wiki/List_of_most-visited_websites

Liu, Q., Xu, K., Zhang, L., Wang, H., Yu, Y., & Pan, Y. (2008). Catriple: extracting triples from Wikipedia categories. *The Semantic Web*, 5367. https://link.springer.com/chapter/10.1007/978-3-540-89704-0_23

Luyt, B., Ally, Y., Low, N. H., & Ismail, N. B. (2010). Librarian perception of Wikipedia: threats or opportunities for librarianship? *Libri, 60,* 57-64. http://late-dpedago.urv.cat/site_media/papers/libr.pdf

Marek, M., Rensing, C. and Steinmetz, R. (2007). *Categorizing learning objects based on Wikipedia as substitute corpus.* https://www.researchgate.net/publication/221549919_Categorizing_Learning_Objects_Based_On_Wikipedia_as_Substitute_Corpus

Medeiros, J. F., Nunes, B. P., Siqueira, S. W. M., & Leme, L. A. P. (2018). *Tag the web: using Wikipedia categories to automatically categorize resources on the web.* https://2018.eswc-conferences.org/files/posters-demos/paper_275.pdf

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39–41.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.1823&rep=rep1&type=pdf

Milne, D., Medelyan, O., & Witten, I. H. (2006). Mining domain specific thesauri from Wikipedia: a case study. *IEEE/WIC/ACM International Conference on Web Intelligence* (pp.442-448). http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4061409&isnumber=4061322

Nastase, V., Strube, M., Boerschinger, B., Zirn, C. & Elghafari, A. (2010). WikiNet: A very large scale multi-lingual concept network. *Proceedings of the 7th International Conference on Language Resources and Evaluation (pp.*1015-1022). http://www.lrec-conf.org/proceedings/lrec2010/pdf/615_Paper.pdf

Nastase, V., & Strube, M. (2008). Decoding Wikipedia categories for knowledge acquisition. *Proceedings of the National Conference on Artificial Intelligence* (pp.1219-24). https://pdfs.semanticscholar.org/b1cc/d016a32e5df7c9543dbe515937b0004fd9fa.pdf.

Pappu, A. (2009). *Using Wikipedia for hierarchical finer categorization of named entities* [Conference session]. 23rd *Pacific Asia Conference on Language, Information and Computation,* (pp.779–786). https://www.aclweb.org/anthology/Y09-2041.pdf.

Paul, D & Pramanik, P. (2013). India in Wikipedia: a study in comparison with Encyclopedia Britannica. *IASLIC Bulletin, 58*(1), 15-23.

Ponzetto, Simone & Strube, Michael. (2007). Deriving a large scale taxonomy from Wikipedia. In *Proceedings of the National Conference on Artificial Intelligence, 2* (pp.1440-1445). https://cdn.aaai.org/AAAI/2007/AAAI07-228.pdf

Pratapa, S, & Kumar, S. (2014). *Deep (hierarchical) classification of Tweets* [Doctoral dissertation, Indian Institute of Technology, Kharagpur]. https://cse.iitkgp.ac.in/~psvijay/docs/MTP_1.pdf.

Qureshi, M. A., Younus, A., O'riordan, C., & Pasi, G. (2014). *CIRGIRDISCO at Rep Lab 2014 Reputation Dimension Task: using Wikipedia graph structure for classifying the reputation dimension of a Tweet*. http://ceur-ws.org/Vol-1180/CLEF2014wn-Rep-QureshiEt2014.pdf

Ragues, B. (2009). Wiki-philosophizing in a marketplace of ideas: evaluating Wikipedia's entries on seven great minds. *eJournal, 2*(1), 117–158. https://papers.ssrn.com/sol3/Data_Integrity_Notice.cfm?abid=978177

Read, B. (2006). Can Wikipedia ever make the grade? as questions about the accuracy of the anyone-can-edit encyclopedia persist, academics are split on whether to ignore it, or start contributing. *The Chronicle, 53.* http://www.nku.edu/~ocallaghant/courses/601/wikipedia_art3.pdf

Rector, L. M. (2008). Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review, 36*(1), 7-22. http://www.emeraldinsight.com/doi/pdfplus/10.1108/00907320810851998

Salah, A., Gao, C., Suchecki, K., & Scharnhorst, A. (2010). Need to categorize: a comparative look at the categories of the Universal Decimal Classification system (UDC) and Wikipedia. *Leonardo, 45*(1), 84-85. https://www.researchgate.net/publication/271246801_Evolution_of_Classification_Systems

Salah, A., Gao, C., Suchecki, K., & Scharnhorst, A. (2011). *Generating ambiguities: mapping category names of Wikipedia to UDC class numbers.* http://www.networkcultures.org/publications

Santosh, K., Joshi, A., Gupta, M., & Varma, V. (2014). Exploiting Wikipedia categorization for predicting age and gender of blog authors. *CEUR Workshop Proceedings* (pp.33-36). https://www.researchgate.net/publication/286753548_Exploiting_wikipedia_categorization_for_predicting_age_and_gender_of_blog_authors/citation/download.

Scharnhorst, A., Smiraglia, R. P., Guéret, C., & Salah, A. (2016). Knowledge maps of the UDC: uses and use cases. *Knowledge Organization, 43*(8), 641-654. https://www.researchgate.net/publication/316645167

Schonhofen, P. (2012). Identifying document topics using the Wikipedia category network. *Web Intelligence and Agent Systems: An International Journal,* 7, 456-462. https://ieeexplore.ieee.org/document/4061411

Soules, A. (2015). Faculty perception of Wikipedia in the California State University system. *New Library World, 116(*3), 213-226. https://www.emerald.com/insight/content/doi/10.1108/NLW-08-2014-0096/full/html

Soules, A. (2015). Faculty perception of Wikipedia in the California State University system. *New Library World*, *116*(3), 213 – 226. http://www.emeraldinsight.com/doi/abs/10.1108/NLW-08-2014-0096

Steward, B. (2004). Writing a literature review. *British Journal of Occupational Therapy, 67*(11), 495-500. https://10.1177/030802260406701105

Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). Yago: a large ontology from Wikipedia and WordNet. *Journal of Web Semantics, 6*(3), 203-217. http://www.sciencedirect.com/science/article/pii/S1570826808000437

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). YAGO: a core of semantic knowledge. *16th International World Wide Web Conference on World Wide Web (WWW 2007)* (pp. 697-706). Association for Computing Machinery, New York, NY, USA. https://10.1145/1242572.1242667

Suchecki, K., Salah, A., & Gao, C. (2012). Evolution of Wikipedia's category structure. *WSPC/Instruction File.* https://arxiv.org/pdf/1203.0788.pdf.

Szymanski, J., & Duch, W. (2012). Self organizing maps for visualization of categories. *In Neural Information Processing, ICONIP 2012, Lecture Notes in Computer Science.*7663. https://link.springer.com/chapter/10.1007/978-3-642-34475-6_20.

Vargas, J. A. (2007, September 17). On Wikipedia debating 2008 hopefuls' every facet. *Washington Post Staff Writer.* https://www.washingtonpost.com/archive/national/2007/09/17/on-wikipedia-debating-2008-hopefuls-every-facet/435ac52f-b010-470a-93b8-4f7d42467b93/

Voss, J. (2009). Wikipedia as knowledge organization system. *International Cataloguing and Bibliographic Control: Quarterly Bulletin of the IFLA UBCIM Programme, 39*(2), 41-42. https://search.proquest.com/openview/7deb2f5435b182877a1ee7557816261d/1?pq-origsite=gscholar&cbl=60376

Wang, P., Hu, J., Zeng, H., & Chen, Z. (2009). Using Wikipedia knowledge to improve text classification. *Knowledge and Information Systems, 19,* 265-281. https://www.semanticscholar.org/paper/Using-Wikipedia-knowledge-to-improve-text-WanHu/fc1d23d2f9167d13ef1bce098ef55d1b40894dd4

Wikipedia: Categorization. (2024, October, 06). In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Wikipedia:Categorization&action=info

Wikipedia: Statistics. (2024, September 10). In *Wikipedia.* https://en.wikipedia.org/w/index.php?title=Wikipedia:Statistics&action=info

Wikipedia. (2024, November 1). In *Wikipedia.* https://en.wikipedia.org/wiki/Wikipedia

Witzleb, N. (2009). Engaging with the world: students of comparative law write for Wikipedia. *Legal Education Review, 19*(1). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1725107

Xiao, L., & Askin, N. (2012). Wikipedia for academic publishing: advantages and challenges. *Online Information Review, 36*(3), 359-357. https://www.researchgate.net/publication/234144804_Wikipedia_for_Academic_Publishing_Advantages_and_Challenges#:~:text=Findings%20%E2%80%93%20Compared%20to%20an%20open,a%20wider%20variety%20of%20articles