

WORDNET DOMAINS 2.0

Pamela Forner
ITC-irst, Povo-Trento, Italy
February 2005
Document Version:1.0

Although the first version of the WDH has been successfully applied to various tasks in different scenarios, it presented some problems mostly related to the lack of a clear semantics of the domain labels. A revision of the WORDNET DOMAINS Hierarchy has been carried out with the aim of providing it with a defined semantics and evaluating coverage and balancing of the *Basic Domains* (the second level of WORDNET DOMAINS Hierarchy), in order to improve the applicability of the WDH in a wider range of fields. The revision is crucially based on the Dewey Decimal Classification (Edition 21) which has been used as a reference to address such problems.

The semantics of the WDH labels has been defined by creating one or more pointers to the DDC codes. The *Basic Domain* coverage has been evaluated by checking that each DDC code is covered by at least one *Basic Domain*. Balancing has been evaluated by making a comparison between the granularity of the *Basic Domains* and that of the *Hundred Division* of the DDC.

The tables below illustrate the changes which have been done with respect to the previous version, followed by a short explanation.

OLD_HIERARCHY

<i>Basic Domain</i>	
Archaeology	
Astrology	
History	
	Heraldry
Linguistics	
	Grammar
Literature	
	Philology

NEW_HIERARCHY

<i>Basic Domain</i>		<i>DDC</i>
History		[920:990]
	Archaeology	930.1
	Heraldry	[929.6,929.7]
Linguistics		410
	Grammar	415
Literature		[800,400]
	Philology	400

ARCHAEOLOGY: as the number of publications can confirm, it is too small a domain to be *basic*. For this reason and consistently with the DDC we have made **ARCHAEOLOGY** (930.1) a sub-domain of **HISTORY** [920:990].

ASTROLOGY: in the DDC it belongs to “Paranormal Phenomena” (130). It is true that it may have a great impact on society but its vocabulary is narrow and the number of publications is limited, so we decided to remove it from the list of the *Basic Domains* given the fact that it is strictly connected with **OCCULTISM** (which was erroneously placed under **RELIGION**), , we have created a new *Basic Domain* **PARANORMAL** (130) which includes **ASTROLOGY** (133.5) and **OCCULTISM** [133-133.5], consistently with the DDC.

PHILOLOGY: in the DDC it has not its own status, but it is classified under a generic “Language” (400). In the old hierarchy **PHILOLOGY** had been put under **LITERATURE** and there is a reason for that: it reflects a label of the Garzanti dictionary’s nomenclature which refers to the reconstruction and correct interpretation of literary documents in a particular culture. With such a semantics **PHILOLOGY** can rightly be placed under **LITERATURE**. The part of **Philology** which is more strictly related to **Linguistics** should obviously be classified under **LINGUISTICS**.

OLD_HIERARCHY

<i>Basic</i>		
Art		
	Dance	
	Drawing	
		Painting
		Philately
	Music	
	Photography	
	Plastic_Arts	
		Jewellery
		Numismatics
		Sculpture
	Theatre	

NEW_HIERARCHY

<i>Basic</i>			<i>DDC</i>
Art			[700- (710,720,745.5,790- (791.43,792,793.3))]
	Graphic_Arts		760
		Philately	769.56
	Dance		[792.8,793.3]
	Drawing		[740-745.5]
	Painting		750
	Music		780
	Photography		770
	Plastic_Arts		730
		Jewellery	739.27
		Numismatics	737
		Sculpture	[731:735]
	Theatre		[792-792.8]
	Cinema		791.43

GRAPHIC_ARTS (760): the old hierarchy did not comprise GRAPHIC_ARTS which have been introduced in the new hierarchy in order to include, consistently with the DDC, PHILATELY (769.56) which previously was inexactly classified under DRAWING [740-745.5].

PAINTING (750): for the sake of consistency with the DDC, PAINTING is no more under DRAWING, but it has become one of its brothers.

CINEMA (791.43): has been moved from TELECOMMUNICATION to ART. Documents concerning cinematography or technical aspects related to telecommunications are to be classified in the general part of TELECOMMUNICATION.

In the DDC **THEATRE** [792-792.8], **DANCE** ([792.8,793.3]) and **CINEMA** (791.43), are not found directly under “The Arts” (700) but are placed under “Recreational and Performing arts” (790) together with “Play” and “Sport”. The DDC makes no distinction between “Art” and “Entertainment” in the sense that “Entertainment” is a part of “Art”. Instead, we make a distinction between the two domains, as FREE_TIME is a Top-level and it is not placed under ART. We have decided to keep THEATRE, DANCE and CINEMA under ART as we think it is the best place, and we remain nevertheless consistent with DDC as there is a connection between “Art” and “Free Time”.

OLD_HIERARCHY

	<i>Basic</i>	
		Theatre
	Religion	
		Mythology
		Occultism
		Roman_Catholic
		Theology
Free_Time		
	Play	

NEW_HIERARCHY

	<i>Basic</i>		<i>DDC</i>
	Paranormal		130
		Occultism	[133-133.5]
		Astrology	133.5
	Religion		200
		Theology	[220,230,240,260]
		Roman_Catholic	282
		Mythology	291.13
Free_Time			[790-(791.43,792,793.3)]
	Radio-Tv		[791.44,791.45]

PARANORMAL: it is a newly created *Basic Domain* (in the DDC it is named “Paranormal Phenomena”, 130) and it contains ASTROLOGY (133.5) and OCCULTISM [133-133.5]. To follow the DDC, we should have made ASTROLOGY a sub-domain of OCCULTISM, but we have preferred to keep them at the same level (as brothers), as we did not want to demote ASTROLOGY too much (it was a *Basic Domain* in the old hierarchy)

RELIGION: it loses OCCULTISM (which has been moved under PARANORMAL); such replacement is perfectly in keeping with the DDC hierarchical structure.

Consistently with the DDC, MYTHOLOGY (291.13) remains under RELIGION (200). The part of Mythology which has nothing to do with Religion is to be classified in FOLKLORE (398)

RADIO AND TV [791.44,791.45]: in the old hierarchy they were two separate domains which have been moved from TELECOMMUNICATION and now they form together a new *Basic Domain* related to FREE_TIME. We have decided to keep them together as they have strict connections. Documents talking about radio and tv technologies are to be categorized in the general part of TELECOMMUNICATION.

OLD_HIERARCHY

<i>Basic</i>	
Agriculture	
Alimentation	
	Gastronomy
Veterinary	
	Zootechnics

NEW_HIERARCHY

<i>Basic</i>			<i>DDC</i>
Agriculture			[630,338.1]
	Animal_Husbandry		[636,638,639]
		Veterinary	636.089
Food			[613.2,613.3,641,642]
	Gastronomy		[642,641.5:641.8-641,62]
Home			[640-(641,642,645)]

In the old hierarchy **VETERINARY** was a *Basic Domain* having **ZOOTECHNICS** as sub-domain, but this state of things finds no correspondence in the DDC where the contrary happens (in the sense that “Veterinary” is the small sub-domain).

As far as **ZOOTECHNICS** is concerned, we have decided to enlarge the semantics of the domain and to change its name to **ANIMAL_HUSBANDRY** [636,638,639], a more generic domain which was missing in the previous hierarchy. Consistently with the DDC, **VETERINARY** (636.089) becomes a sub-domain of **ANIMAL_HUSBANDRY** [636,638,639] which finds its place under **AGRICULTURE** [630,338.1]. The semantics of the label Agriculture is the same as in the DDC, i.d. “Agriculture and related technologies”, of which **ANIMAL_HUSBANDRY** is an integral part. The position of **VETERINARY** remains an open issue, as it regards not only reared animals but also pets and domestic animals.

ALIMENTATION: (“Food & drink” in the DDC) [613.2,613.3,641,642]: to be consistent with the DDC we have substituted the name of the discipline Alimentation with the name of the semantic field “**Food**”. The change of terminology has also been conceived in order to reflect current and international usage (see Google directories).

HOME [640-(641,642,645)]: it is a newly created *Basic Domain* which was missing in the old hierarchy and has been created to include the topics related to home economics and family living (for example: housekeeping, household utilities, etc.)

OLD HIERARCHY

<i>Basic</i>	
Architecture	
	Town_Planning
	Building_Industry
	Furniture
Engineering	
	Mechanics
	Astronautics
	Electrotechnics
	Hydraulics
Astronomy	
	Topography

NEW HIERARCHY

<i>Basic</i>		<i>DDC</i>
Architecture		[645,690,710,720]
	Town_Planning	710
	Buildings	690
	Furniture	645
Engineering		620
	Mechanics	620.1
	Astronautics	629.4
	Electrotechnology	621.3
	Hydraulics	627
Telecommunication		[383,384]
	Post	383
	Telegraphy	384.1
	Telephony	384.6
Astronomy		520

ARCHITECTURE: in the DDC, both “Architecture” (720) and “Town_Planning” (710) are placed under “Art” (700), while “Furniture” (645) is contained in the section “ Home economics and family living” (640). Supported by the view of experts in the filed , we have not modified the hierarchy so that ARCHITECTURE remains a *Basic Domain* (as it was in the old hierarchy) and also all its sub-domains are preserved.

ELECTROTECHNICS: it has no code in the DDC . Native speakers do not recognize the word and are not able to explain the meaning; Collins and Sansoni dictionaries give electrotechnology as translation equivalent. The term also finds a confirmation on the web (see Google): from a search of the word, electrotechnology results to be a discipline of engineering. We have changed the name of the label to electrotechnology (but even electrotechnology has no code in the DDC).

TELECOMMUNICATION: we have changed its position in the hierarchy. It is not placed under SOCIAL_SCIENCE anymore, but it is located under APPLIED_SCIENCE. Moreover, it loses RADIO, CINEMA, TELEVISION (which have been moved under ART and FREE_TIME). It goes without saying that all technological aspects of the mentioned domains are to be classified in the general part of such domain.

ASTRONOMY: it loses TOPOGRAPHY as, consistently with the DDC, it has been moved under GEOGRAPHY.

OLD HIERARCHY

<i>Basic</i>		
Biology		
	Biochemistry	
	Ecology	
	Botany	
	Zoology	
		Entomology
	Anatomy	
	Physiology	
	Genetics	

NEW HIERARCHY

<i>Basic</i>		<i>DDC</i>
Biology		[570-577,611,612-612.6]
	Biochemistry	572
	Anatomy	611
	Physiology	[571,612,-612.6]
	Genetics	576
Animals		590
	Entomology	595.7
Plants		580
Environment		577

BIOLOGY: To be consistent with the DDC we have changed the name of the labels:

ZOOLOGY has become **ANIMALS** (590)

BOTANY has become **PLANTS** (580)

Moreover these two domains are no more under BIOLOGY [570-577,611,612-612.6], but they have become *Basic Domains*. Also this change is consistent with the DDC where they are brothers of “Biology” (570).

We have decided to change also the name of the label ECOLOGY which has become **ENVIRONMENT**. This substitution is due to the fact that we give an enlarged interpretation to the semantics of the label which is more connected with environmental matters. The meaning of the label embraces also part of the human ecology (see the definition of ecology in “The American Heritage® Dictionary of the English Language”)¹. If we intend the label in this way, Ecology/Environment becomes a subject of great impact on society and it has the right to become a *Basic Domain* (in contrast with the DDC where it is a sub-domain of “Biology” (570)). In the DDC there is no code for Ecology with such an enlarged semantics.

¹ Ecology **1a.** The science of the relationships between organisms and their environments. Also called *bionomics*. **b.** The relationship between organisms and their environment. **2.** The branch of sociology that is concerned with studying the relationships between human groups and their physical and social environments. Also called *human ecology*. **3.** The study of the detrimental effects of modern civilization on the environment, with a view toward prevention or reversal through conservation. Also called *human ecology*.

OLD HIERARCHY

<i>Basic</i>	
Earth	
	Geology
	Meteorology
	Oceanography
	Paleontology
	Geography

NEW HIERARCHY

<i>Basic</i>			<i>DDC</i>
Earth			[550,560,910-(910.4,910.202)]
	Geology		[551-(551.5,551.6,551.46)]
	Meteorology		551.5,551.6
	Oceanography		551.46
	Paleontology		560
	Geography		[910-(910.4,910.202)]
		Topography	912

EARTH: according to the DDC , “Paleontology” (560) and “Geography” [910-(910.4,910.202)] do not belong in “Earth sciences ” (550). “Paleontology” is a brother and “Geography” is classified in another section. We have not modified the old hierarchy as:

- in the Garzanti dictionary’s nomenclature, PALEONTOLOGY is kept together with GEOLOGY (“Geology” is to be found under “Earth sciences” (550))
- the label of GEOGRAPHY has a broad meaning which includes physical, human, economic and political aspects , but this is not enough to make it a *Basic Domain*. Documents regarding physical aspects of Geography are to be classified with good reason here. In our hierarchy there is no other better position to place it , but the question remains still under discussion.

In the DDC METEOROLOGY and OCEANOGRAPHY are sub -domains of “Geology” (not brothers as in our hierarchy). The old hierarchy has been maintained unchanged

TOPOGRAPHY: has been moved from ASTRONOMY to GEOGRAPHY, consistently with the DDC.

OLD HIERARCHY

<i>Basic</i>		
Anthropology		
	Ethnology	
		Folklore
Body_Care		

NEW HIERARCHY

<i>Basic</i>			<i>DDC</i>
Anthropology			301:307,395,398
	Ethnology		305.8
		Folklore	398
Health			[613- (613.2,613.3,613.8,613.9)]
	Body_Care		613.4

ANTHROPOLOGY: this hierarchy is not in keeping with the DDC where no distinction is made between Anthropology and Sociology as they are classified together (301).

ETHNOLOGY (305.8): in the DDC, it is not a sub-domain of ANTHROPOLOGY but it is found under “Social groups” (305)

FOLKLORE (398): in the DDC it is inserted in another section; it is a sub -domain of “Customs, etiquette, folklore” (390).

Anyway all the three domains are contained in “Social Sciences” (300).

We have not altered the old hierarchy as, although it is true that Sociology and Anthropology have something in common, it is also true that they are interested in different matters, so it is not an error to leave them separate. Another suggestion could be:

sociology
 anthropology
 ethnology
 folklore

but the idea of making ANTHROPOLOGY a sub-domain of SOCIOLOGY, was not convincing. Also the definition of the dictionaries seems to support our hierarchy, while the change of position of Anthropology under Sociology appears incorrect.

HEALTH: in the old hierarchy a domain was missing which incorporated the subjects regarding health in general. For this reason we have created a new *Basic Domain* which assembles the topics about BODY_CARE (613.4) (clothing and cosmetics) and others, like for example physical fitness, relaxation, yoga, et c. In the DDC the discipline “Promotion of health” (613) is classified under “Medical Sciences” (610), but as there is plenty of documents talking about Health we have preferred to make it a *Basic Domain*.

OLD HIERARCHY

<i>Basic</i>	
Industry	
	Textiles
Transport	
	Aeronautic
	Auto
	Merchant_Navy
	Railway

NEW HIERARCHY

<i>Basic</i>		<i>DDC</i>
Industry		[338-(338.1,338.642),660,670,680]
Transport		[385:389]
	Aviation	[387.7,387.8]
	Vehicles	[388-388.1]
	Nautical	[386,387.1:387.5]
	Railway	385

INDUSTRY: in the old hierarchy it had only one sub-domain TEXTILE, which has been eliminated (only two synsets in WordNet).

TRANSPORT: the hierarchy has not been modified, only the name of the sub-domains have been changed as they were not homogeneous in the sense that some of them were too generic and others were too specific:

Aeronautics is a discipline of Engineering: it deals with the technology of designing, building and operating aircrafts (also in the DDC it is classified under "Engineering").

Merchant Navy seems to be restricted to the trading of goods.

Auto is too narrow and specific, it does not include all other means of ground transportation, urban transportation, etc.

We have changed only the name of the labels to give them a more general meaning in order to reflect the semantics of the domain, but the hierarchy has been maintained.

- **AVIATION:** everything that has to do with airplanes and air transportation in general (documents/synsets dealing with aeronautics in a technical sense are to be classified in Engineering)
- **RAILWAY:** everything that has to do with trains and rail transportation in general
- **NAUTICAL:** everything that has to do with ships, vessels, boats sea and inland transportation in general
- **VEHICLES** everything that has to do with cars, lorries, motorcycles, bus and ground transportation in general.

In the DDC the terminology is more centered on the transportation aspects, we prefer, instead, our domain to include also what has to do with the means of transportation.

OLD HIERARCHY

<i>Basic</i>	
Economy	
	Banking
	Book_Keeping
	Enterprise
	Exchange
	Insurance
	Money
	Tax

NEW HIERARCHY

<i>Basic</i>			<i>DDC</i>
Economy			[330-338,368,650]
	Enterprise		650
		Book_Keeping	657
	Finance		[332,336.3]
		Banking	[332.1:332.3]
		Money	332.4
		Exchange	332.6
	Insurance		368
	Tax		[336.1,336.2]

ECONOMY: we have reorganized the domain in keeping with the DDC. In order to be completely consistent with the DDC the following changes should be added:

- ENTERPRISE (with its sub -domain BOOK_KEEPING) should NOT be under ECONOMY but in a different section
- INSURANCE should NOT be under ECONOMY but in a different section
- TAX is a many-sided subject which only in minimal part is under ECONOMY
- INDUSTRY in our hierarchy is a *Basic Domain*, but in the DDC it is classified under “Economy”

We have not done these changes as ENTERPRISE and INSURANCE are not big enough to be *Basic Domains* and we prefer not to move TAX from ECONOMY as it has deep connections with it.

PSYCHOLOGICAL_FEATURES: we have created a new domain under **FACTOTUM** where generic synsets/documents dealing with traits of personality, dispositions, emotions and feelings are to be classified. The synsets have been moved from the psychology domain as they do not have a specific usage.

To be done in the future:

1. Proposal for the organization of the top level trying to be as much as possible consistent with the DDC . Suggestions: economic sciences, biological sciences (or life sciences => DDC)
2. creation of a new basic domain (MEDIA) which has as subdomains radio, TV, and internet.
3. search for a better position for veterinary
4. search for a better position for geography
5. study of the hierarchy of anthropology considering its placement under sociology
6. study of the possibility to move artisanship under industry (or creation of a new basic domain “production” or “production activities” where both artisanship and industry can be located, but...agriculture and tourism...should be moved too?)
7. study of the possibility to move commerce under economy... (but commerce is a production activity like industry and agriculture...)