

Frank Busse

DDC-Kurznotationen

Entwicklung & Maschinelle Klassifikation

Inhaltsverzeichnis

1. Maschinelle Klassifikation
2. DDC-Kurznotation
3. Entwicklungsprozess
4. Kennzeichnung
5. Ergebnisse
6. Herausforderungen & Ausblick

Maschinelle Klassifikation

Überwachtes maschinelles Lernen

- Lernen aus Beispielen
- Erkennen von Mustern
- Verallgemeinerung der Muster
- Unbekannte Objekte können klassifiziert werden

Anwendung in der DNB

Software: Averbis Extraction Platform / [Averbis GmbH](#)

Klassifikator: Support Vector Machine (SVM)

Anwendungsgebiete:

- Maschinelle Vergabe von DDC-Sachgruppen
 - Reihe O, Reihe B, Reihe H
- Maschinelle Vergabe von DDC-Kurznotationen
 - Ausgewählte SG der Reihe O,
 - SG 610 Reihe O, Reihe B und Reihe H

Workflow

Training

- Trainingsmaterial
- Modellerstellung
- Software:
 - Averbissoftware



Routine

- Täglicher Zugang
- Retro-Läufe
- Software:
 - Averbissoftware
 - DNB Schnittstelle
 - CBS

DDC-Kurznotationen

Was ist eine DDC-Kurznotation

- DDC-Kurznotationen sind Teil des Gesamtsystems der DDC
- DDC-Kurznotationen sind ein fachlich ausgewähltes und feststehendes Set an DDC-Notationen
- Ausgangspunkt für die Entwicklung der DDC-Kurznotationen ist die DDC-Kurzausgabe (Abridged DDC, Ausgabe 15)
- Bei ihrer Auswahl wird das Literaturlaufkommen in der DNB der letzten zehn Jahre berücksichtigt
- Die Notationslänge und damit der Spezifizierungsgrad weicht deshalb häufig von der Kurzausgabe ab

DDC-Kurznotationen: Ursprung

- Ursprünglich entwickelt 2005/2006 für die Sachgruppe Medizin
- Nur für die Erschließung gedruckter medizinischer Dissertationen (SG 610 Reihe H)
- Seit 2015 maschinelle Vergabe von medizinischen Kurznotationen für Netzpublikationen
- Seit 2017 Ausweitung auf weitere Sachgruppen

Wie sieht eine Kurznotation aus?

DDC-SG	610
DDC vollständig	618.92398009435123090511
Kurznotation	618.92

DNB-SG	004
DDC vollständig	005.82
Kurznotation	005.82

DDC-SG	300
DDC vollständig	303.6250882970956
Kurznotation	303.6

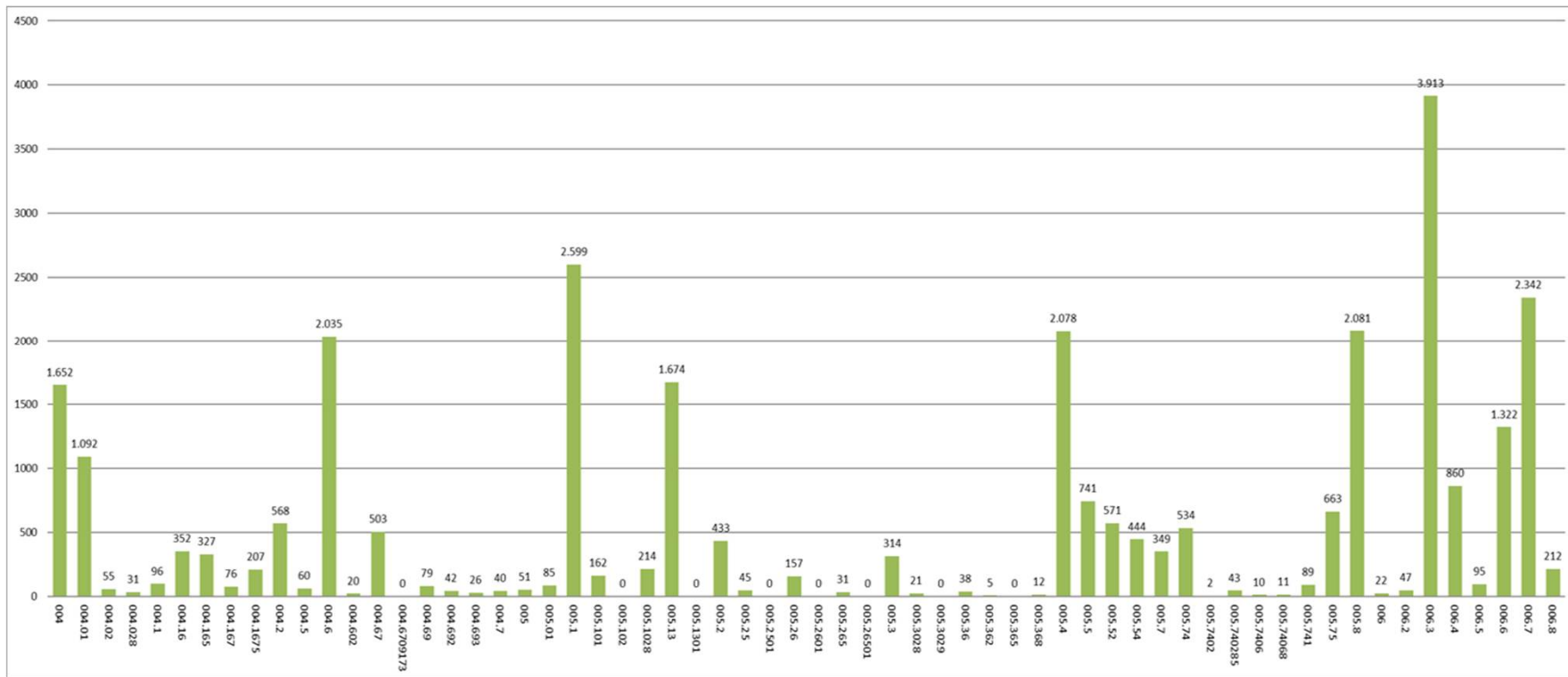
DDC-SG	640
DDC vollständig	641.563620954165
Kurznotation	641.56362

Entwicklungsprozess

DDC-Kurznotationen: Erstellungsprozess

- Schritt I : Ausgangspunkt DDC Abridged Edition 15
- Schritt II : Analyse Literaturaufkommen

Schritt II: Analyse Literaturaufkommen



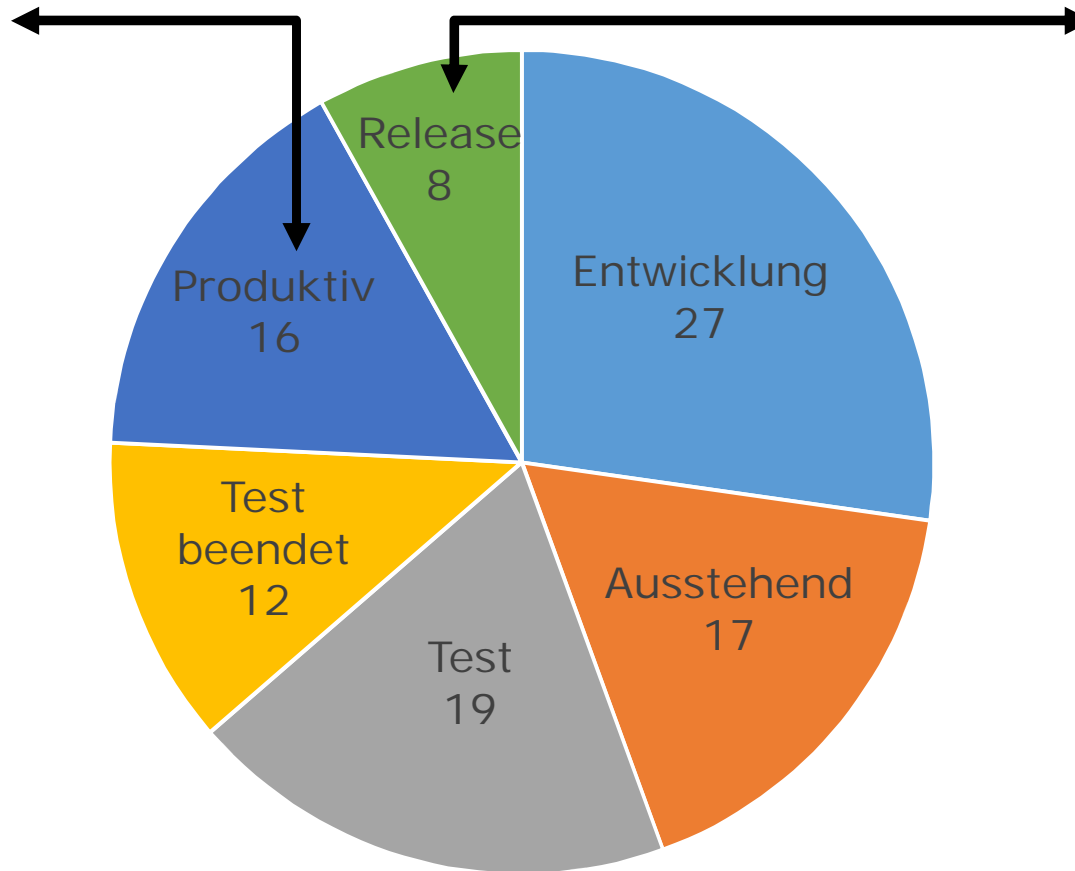
DDC-Kurznotationen: Erstellungsprozess

- Schritt I : Ausgangspunkt DDC Abridged Edition 15
- Schritt II : Analyse Literaturlaufkommen
- Schritt III : Anpassen
 - Entfernen von „unnötigen“ Notationen
 - Identifizierung von weiteren notwendigen Notationen
- Schritt IV : Testen
 - Testen-analysieren-anpassen, testen-analysieren-anpassen
- Schritt V : Produktiver Betrieb

Entwicklungsstand Kurznotationen

SG Produktiv:

- 004
- 300
- 500
- 510
- 520
- 530
- 540
- 600
- 610
- 621.3
- 624
- 640
- 650
- 660
- 670
- 690



Release:

- 130
- 550
- 560
- 570
- 580
- 590
- 620
- 630

Kennzeichnung

Kennzeichnung

- Kennzeichnung aller maschinell vergebenen Kurznotationen
- Datenauslieferung in MARC 21 seit Mitte Mai 2018
- Kennzeichnung und Anzeige im DNB Portal

Kennzeichnung DNB Portal

Link zu diesem Datensatz	http://d-nb.info/1127024027
Titel	Insomnia: Medical Sleep Disorder & Diagnosis / Md Belal Bin Heyat
Person(en)	Heyat, Md Belal Bin (Verfasser)
Ausgabe	1. Auflage
Verlag	Hamburg : Anchor Academic Publishing
Zeitliche Einordnung	Erscheinungsdatum: 2017
Umfang/Format	Online-Ressourcen, 56 Seiten (pdf)
Andere Ausgabe(n)	Elektronische Reproduktion: ISBN: 9783960675891
Persistent Identifier	URN: urn:nbn:de:101:1-2017030745
URL	http://www.anchor-publishing.com/e-book/337729/insomnia-medical-sleep-disorder-diagnosis (Verlag)
ISBN/Einband/Preis	978-3-96067-089-6
EAN	9783960670896
Sprache(n)	Englisch (eng)
Anmerkungen	Lizenzpflichtig. - Vom Verlag als Druckwerk on demand und/oder als E-Book angeboten Langzeitarchivierung gewährleistet
DDC-Notation	616.8 (maschinell ermittelte DDC-Kurznotation)
Sachgruppe(n)	610 Medizin, Gesundheit

Kennzeichnung DNB Portal

Link zu diesem Datensatz	http://d-nb.info/1127024027	
Titel	Insomnia: Medical Sleep Disorder & Diagnosis / Md Belal Bin Heyat	
Person(en)	Heyat, Md Belal Bin (Verfasser)	
Ausgabe	1. Auflage	
Verlag	Anchor Publishing	
Zeitlich	DDC-Notation	616.8 (maschinell ermittelte DDC-Kurznotation)
Umfang	Sachgruppe(n)	610 Medizin, Gesundheit
Andere		
Persist		
URL	http://www.anchor-publishing.com/e-book/337729/insomnia-medical-sleep-disorder-diagnosis (Verlag)	
ISBN/Einband/Preis	978-3-96067-089-6	
EAN	9783960670896	
Sprache(n)	Englisch (eng)	
Anmerkungen	Lizenzpflichtig. - Vom Verlag als Druckwerk on demand und/oder als E-Book angeboten Langzeitarchivierung gewährleistet	
DDC-Notation	616.8 (maschinell ermittelte DDC-Kurznotation)	
Sachgruppe(n)	610 Medizin, Gesundheit	

Kennzeichnung Marc 21

```

XXXXXnam a22XXXXXuc 4500
001 1127024027
003 DE-101
005 20180412100712.0
007 cr|||||||||
008 170307s2017 gw |||||o||| 00|||leng
015 $a17_004$2dnb
016 7 $2DE-101$a1127024027
020 $a9783960670896$9978-3-96067-089-6
024 3 $a9783960670896
024 7 $2urn$aurm:nbn:de:101:1-2017030745
035 $(DE-599)DNB1127024027
040 $a1240$bger$cDE-101$d1247
041 $aeng
044 $cXA-DE-HH
082 74$84p$a616.8$qDE-101$223kdnb
083 7 $a610$qDE-101$223sdnb
100 1 $0(DE-588)1127043552$0http://d-nb.info/gnd/1127043552$(DE-101)1127043552$aHeyat, Md Belal Bin$eVerfasser$4aut
245 00$aInsomnia: Medical Sleep Disorder & Diagnosis$cMd Belal Bin Heyat
250 $a1. Auflage
259 $a11
264 1$aHamburg$bAnchor Academic Publishing$c2017
300 $aOnline-Ressourcen, 56 Seiten
336 $aText$bbt$2rdacontent
337 $aComputermedien$bc$2rdamedia
338 $aOnline-Ressource$bcr$2rdacarrier
500 $aLizenzpflichtig. - Vom Verlag als Druckwerk on demand und/oder als E-Book angeboten
583 1 $aLangzeitarchivierung gewährleistet$ILZA
650 7$81p$(DE-588)4025013-1$0http://d-nb.info/gnd/4025013-1$(DE-101)04025013X$aHirnkrankheit$2gnd
650 7$82p$(DE-588)4171595-0$0http://d-nb.info/gnd/4171595-0$(DE-101)041715950$aNeuropsychiatrie$2gnd
650 7$83p$(DE-588)1068493003$0http://d-nb.info/gnd/1068493003$(DE-101)1068493003$aNervenkrankheit$2gnd
653 $a(Produktform)Electronic book text
653 $a(BISAC Subject Heading)TEC007000
653 $aInsomnia;Power Spectral Density;Diagnosis;Sleep Disorder;Short Time Frequency;EEG Signal
653 $a(VLB-WN)1684
776 08$IElektronische Reproduktion$z9783960675891
850 $aDE-101a$aDE-101b
856 40$uhttp://nbn-resolving.de/urn:nbn:de:101:1-2017030745$xResolving-System
856 0$uhttp://d-nb.info/1127024027/34$xLangzeitarchivierung Nationalbibliothek
856 4 $qapplication/pdf$uhttp://www.anchor-publishing.com/e-book/337729/insomnia-medical-sleep-disorder-diagnosis$xVerlag
883 1 $81p$amaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 1 $82p$amaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 1 $83p$amaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 0 $84p$amaschinell gebildet$d20170307$qDE-101
925 r $aro$aara
925 p $apd
  
```

Kennzeichnung Marc 21

```
XXXXnam a22XXXXuc 4500
001 1127024027
003 DE-101
005 20180412100712.0
007 cr|||||
008 170307s2017 gw |||||o||| 00|||leng
015 $a17,004$d2dnb
016 7 $2DE-101$a1127024027
020 $a9783960670896$9978-3-96067-089-6
024 3 $a9783960670896
024 7 $2urn$aurm:nbn:de:101:1-2017030745
035 $(DE-599)DNB1127024027
040 $a1240$bger$cDE-101$d1247
041 $aeng
044 $cXA-DE-HH
082 74$84lp$a616.8$qDE-101$223kdnb
083 7 $a610$qDE-101$223sdnb
100 1 $(DE-588)1127043552$0http://d-nb.info/gnd/1127043552$(DE-101)1127043552$aHeyat, Md Belal Bin$eVerfasser$4aut
245 00$aInsomnia: Medical Sleep Disorder & Diagnosis$cMd Belal Bin Heyat
250 $a1. Auflage
259 $a11
264 1$aHamburg$bAnchor Academic Publishing$c2017
300 $aOnline-Ressourcen, 56 Seiten
336 $aText$btxt$2rdacontent
337 $aComputermedien$bcr$2rdamedia
338 $aOnline-Ressource$bcr$2rdacarrier
500 $aLizenzpflichtig. - Vom Verlag als Druckwerk on demand und/oder als E-Book angeboten
583 1 $aLangzeitarchivierung gewährleistet$ILZA
650 7$81p$(DE-588)4025013-1$0http://d-nb.info/gnd/4025013-1$(DE-101)04025013X$aHirnkrankeheit$2gnd
650 7$81p$(DE-588)4025013-1$0http://d-nb.info/gnd/4025013-1$(DE-101)04025013X$aHirnkrankeheit$2gnd
650 7$81p$(DE-588)4025013-1$0http://d-nb.info/gnd/4025013-1$(DE-101)04025013X$aHirnkrankeheit$2gnd
650 7$81p$(DE-588)4025013-1$0http://d-nb.info/gnd/4025013-1$(DE-101)04025013X$aHirnkrankeheit$2gnd
776 08$eElektronische Reproduktion$9783960675891
850 $aDE-101a$aDE-101b
856 40$uhttp://nbn-resolving.org/urn:nbn:de:101:1-2017030745$xResolving-System
856 0$uhttp://d-nb.info/1127024027/345xLangzeitarchivierung Nationalbibliothek
856 4 $qapplication/pdf$uhttp://www.anchor-publishing.com/e-book/337729/insomnia-medical-sleep-disorder-diagnosis$xVerlag
883 1 $81p$aMaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 1 $82p$aMaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 1 $83p$aMaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
883 0 $84p$aMaschinell gebildet$d20170307$qDE-101
883 1 $85p$aMaschinell aus Konkordanz gebildet$c1$d20170316$qDE-101
925 1 $aeng
925 p $apd
```

Notation

Ausgabe

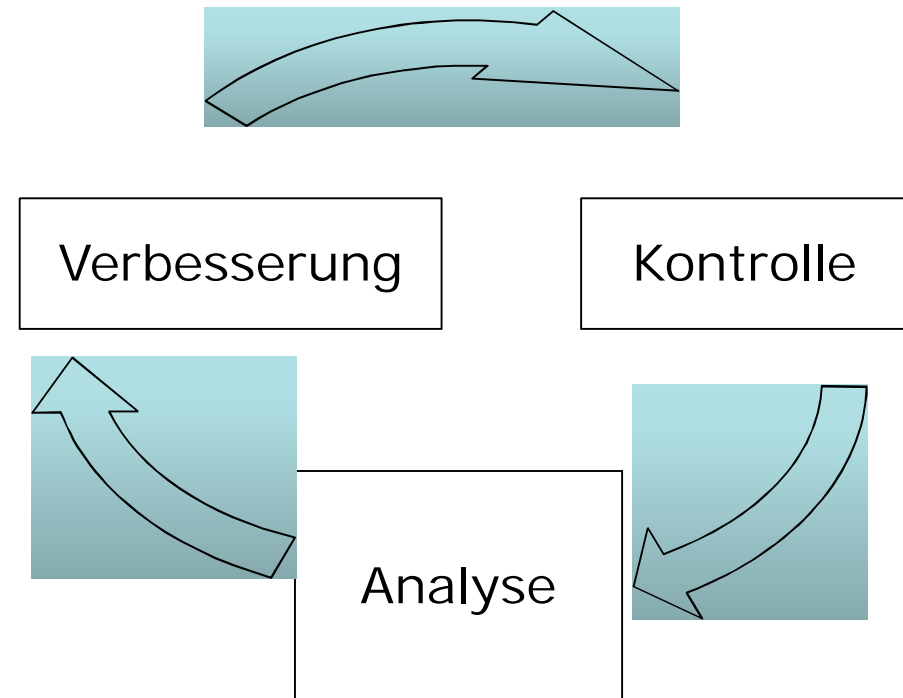
082 74\$84lp\$a616.8\$qDE-101\$223kdnb
083 7 \$a610\$qDE-101\$223sdnb

883 0 \$84p\$aMaschinell gebildet\$d20170307\$qDE-101

Ergebnisse

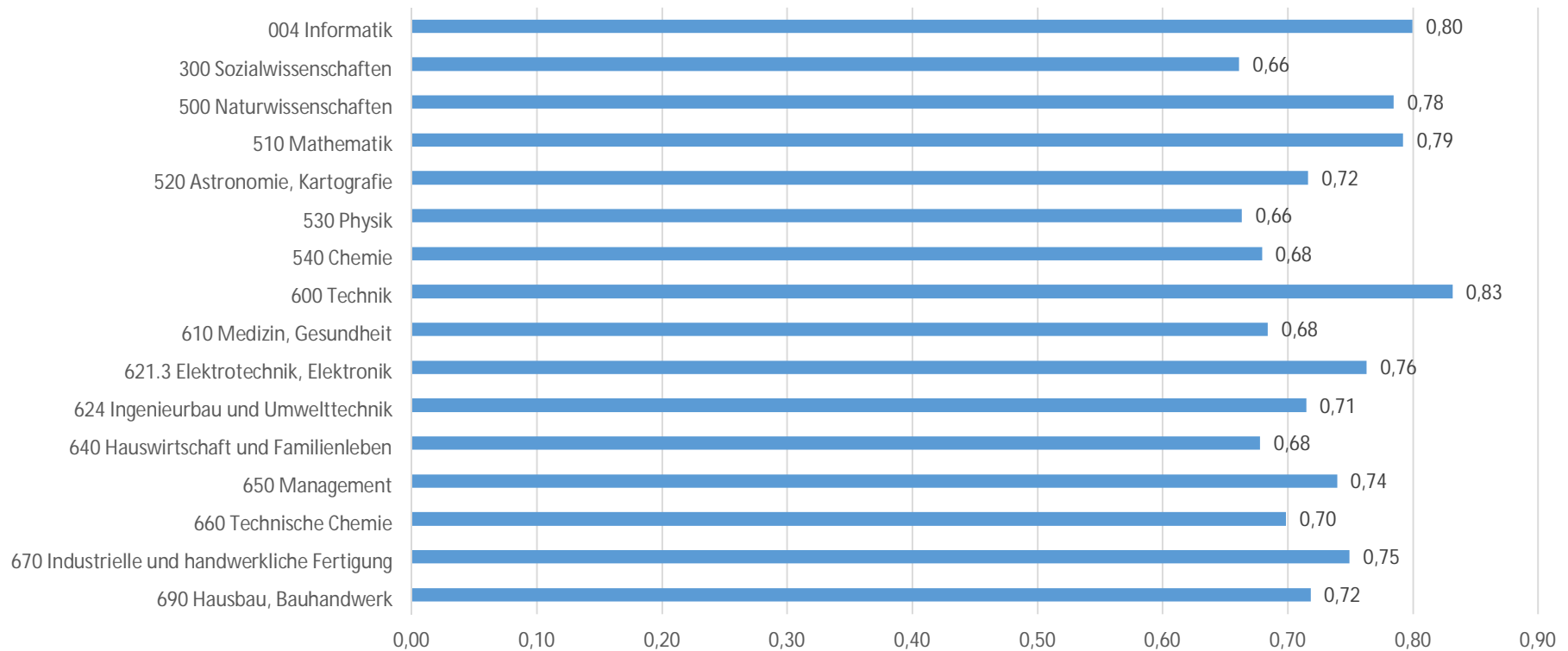
Grundlage

- Übereinstimmung zwischen maschineller und intellektuell vergebener Notation
- Kontrollmöglichkeiten
 - Intellektuelle Überprüfung
 - Parallelausgaben

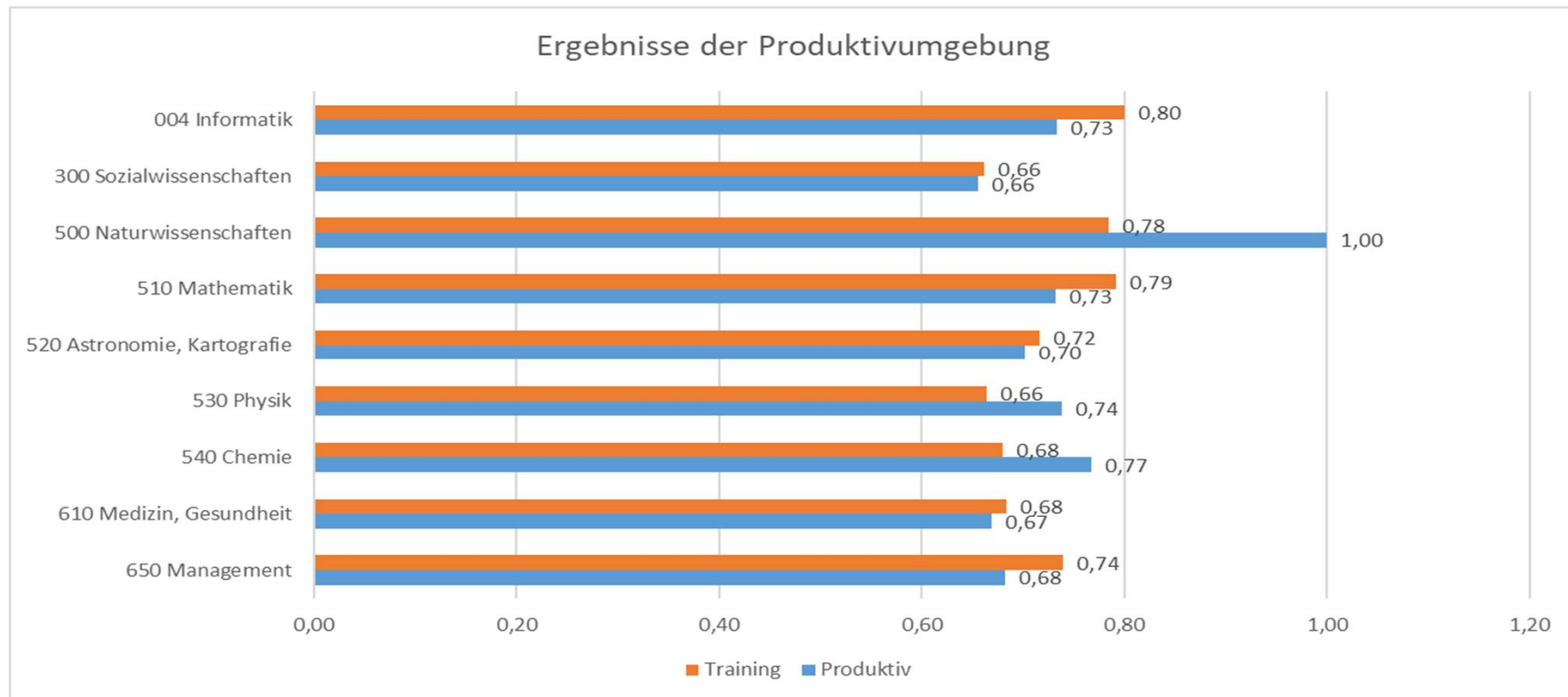


Ergebnisse aus der Testumgebung

F1-Measure Training



Ergebnisse aus der Produktivumgebung 1. Halbjahr 2019



Herausforderungen & Ausblick

Herausforderungen

- Folgefehler im Produktivbetrieb:
Die maschinelle Vergabe der Kurznotationen ist von der maschinell vergebenen DDC-SG abhängig
- Viele kleine Sachgruppen mit wenig Trainingsmaterial
- (Keine) Qualitätsstandards
- Umgang mit Änderungen in der DDC (Neue Themen)

Ausblick

- Verbesserung der Ergebnisse
- DCC-Kurznotationen für alle DDC-Sachgruppen
- Projekt Ema – neue Verfahren

Vielen Dank für Ihre Aufmerksamkeit!

Fragen?

Frank Busse

Deutsche Nationalbibliothek

Automatische Erschließungsverfahren, Netzpublikationen

Telefon: 069-1525-1550

<mailto:f.busse@dnb.de>