
OVERZICHT OPLEVERING CBS 2.1 NOVEMBER 2004

Aan: GEBRUIKERS GGC EN NCC
Van: Daniël van Spanje
cc: Servicedesk OCLC PICA
Datum: 1 november 2004
documentnaam: Overzicht oplevering CBS 2.1 2004
documentstatus: extern
versie: 1.0

INHOUDSOPGAVE

1	INLEIDING	2
2	UNICODE	3
2.1	OPSLAG IN DE DATABASE	3
2.2	INDEXEN EN ZOEKEN	3
2.3	UNICODE EN LBS3	4
2.4	UNICODE EN OUF	4
2.5	UNICODE EN UITVOER	4
3	SET SORTERING	4
4	ZOEKEN MET WOORDCOMBINATIES.....	5
5	IBL.....	5
5.1	VELDEN IN BIBLIOTHEEKADRESSEN 1 REGELIG	5
5.2	BIBLIOTHEEKTYPE "ONBEKEND"	5
5.3	BONNEN EN ANTWOORDCODES VIA E-MAIL	5
5.4	KORTE PRESENTATIE DEPOSITO'S	6
5.5	ONDERLINGE VERREKENING	6
5.6	SELECTEREN VAN IBL-AANVRAGEN.....	6
6	FORMAATUITBREIDINGEN	7
7	STATISTIEK	7
7.1	TOT IN PLAATS VAN T/M	7
7.2	ELEKTRONISCHE AANVRAGEN	7
7.3	TELLERS 201-212.....	7
7.4	TELLER BIJ INLOGGEN	8

1 Inleiding

Deze notitie bevat een overzicht van de aanpassingen en uitbreidingen die door OCLC PICA in versie 2.1. van het centrale systeem worden opgeleverd. Deze uitbreidingen zijn van belang voor gebruikers van het GGC en van het NCC/IBL-systeem.

De notitie wordt verspreid via de Listserver GGC, de discussielijst van het CHUB (Convent Hoofden Uitleenbureaus Universiteitsbibliotheken) en de via de WSF-vertegenwoordiger in dit laatste forum naar de WSF-IBL-functionarissen.

De inhoud van deze notitie wordt ook opgenomen op de website van OCLC PICA.

CBS versie 2.1 is opnieuw een grote systeemoplevering en een volgende belangrijke stap in de invoering van Unicode. Na de voorbereidende stappen in vorige opleveringen wordt nu de mogelijkheid geboden om Unicode-indexen te maken en zullen na de oplevering op 22 november de bibliografische records ook in Unicode-formaat opgeslagen worden in de database.

Voor sommige gebruikers zullen de nieuwe mogelijkheden eerder als complicerend dan als vereenvoudigend ervaren worden. Van belang is de Unicode-functionaliteit met name voor gebruikers die belang hechten aan het invoeren en verwerken van niet-westerse schriften. Maar ook voor gebruikers die publicaties verwerken met formules in titels of gedeelten niet-westers tekst ontstaan verbeterde mogelijkheden.

Dit is echter nog niet de laatste stap. Om over volledige Unicode-functionaliteit te beschikken moeten nog twee stappen gezet worden. Allereerst is voor het invoeren van Unicode-records nog een nieuwe versie van WinIBW nodig. De implementatie van deze nieuwe versie (WinIBW 3.0) is gepland voor het tweede kwartaal van 2005. De andere stap is het "dubbel" kunnen invoeren van originele en getranslitererde titels in één bibliografisch record. De aanpak van de invoering van deze nieuwe functionaliteit en – daarmee samenhangend – de update van transliteratie-mogelijkheden en -tabellen zal besproken worden in de Werkgroep Kwaliteit Databank en de Redactie Richtlijnen.

In deze CBS versie 2.1 wordt voor de IBL-gebruikers een belangrijke stap richting e-mail-implementatie voor de communicatie met het IBL-systeem gezet. Dit betreft het via e-mail laten toesturen van IBL-bonnen, het terugsturen van IBL-antwoordcodes en het naar aanleiding daarvan ontvangen van een "status-or-error-report". Tot nu toe waren die laatste overzichten beperkt tot error-boodschappen. Voor deze functionaliteit is een aparte notitie beschikbaar die door belangstellenden kan worden opgevraagd bij de Servicedesk.

Een derde belangrijke uitbreiding van CBS 2.1 is de uitbreiding van het bibliografische formaat voor de beschrijving van elektronische bronnen - en dan met name voor het beschrijven van e-journals –. Tegelijk met deze uitbreidingen wordt een aantal andere formaat-uitbreidingen ingevoerd, zoals aanpassingen voor invoer van het 13-cijferige isbn.

Een overzicht van de nieuwe kenmerkcodes en het voorstel voor de Richtlijnen voor Online Resources is reeds verspreid via de Listserver GGC.

Tenslotte zijn er nog enkele andere uitbreidingen, zoals de nieuwe GOO-periode-aanduiding, enkele kleinere statistiek- en IBL-aanpassingen en de mogelijkheid om na het zoeken de gevonden set alfabetisch te sorteren. Over de invoering van de nieuwe GOO-periode worden de GOO-gebruikers ook apart geïnformeerd.

2 Unicode

2.1 Opslag in de database

Vanaf de oplevering van CBS-versie 2.1 worden alle records in het GGC, dus zowel bibliografische als thesaurus-records in Unicode opgeslagen. OCLC PICA maakt voor de opslag gebruik van UTF-8, de de facto standaard voor Unicode.

Maar aangezien UTF-8 invoer en bewerkingen pas volledig mogelijk zijn met WinIBW 3.0 worden de records bij gebruik in het GGC in een groot aantal situaties teruggeconverteerd naar de Pica-characterset. Deze Pica-characterset kan globaal gelijk gesteld worden aan de ISO-Latin1-characterset. Dit terugconverteren in een bewerkingsscherm gebeurt indien de client-software, zoals bijvoorbeeld WinIBW versie 2.4.x en ouder, niet Unicode-compliant zijn. Er zijn andere situaties, bijvoorbeeld in titelpresentaties waarbij Unicode-characters die niet gepresenteerd kunnen worden als escape-code worden gepresenteerd ("&#x<nr>", waarbij <nr> staat voor de hexadecimale notatie). In weer andere situaties, bijvoorbeeld bij OUM en OUF, zullen Unicode-characters die niet in de Pica-characterset voorkomen, bij uitvoer worden verwijderd.

Zodra een GGC-record na invoer of mutatie in de database wordt opgeslagen, wordt aan het record een systeemcode toegekend met als inhoud "utf8". Zo kan altijd nagegaan worden of een record al een keer geconverteerd en in UTF8 is opgeslagen. De database zal dus niet in één keer worden geconverteerd, maar de komende jaren langzamerhand tijdens en door gebruik geconverteerd worden naar UTF-8.

De nieuwe kmc's zijn:

	Pica3 code (titels)	Pica3 code (ingangen)	Pica+ code
Algemeen nivo	000K	00K	001U \$0
Lokaal nivo	0247	00L	101U \$0
Exemplaar nivo	0248	00M	201U \$0

2.2 Indexen en zoeken

Met CBS-versie 2.1 kan ook in de indexen en bij het zoeken gebruik gemaakt worden van Unicode. De nieuwe Unicode versie van de zoek- en indexengine (PSI) is daartoe in het CBS ingebouwd. Alle inductabellen zijn opnieuw opgebouwd waarbij uitgegaan is van de bestaande indexdefinities. Er zijn dus geen extra indexen aangemaakt. De oude indexen worden wel op nieuwe Unicode wijze aangemaakt. Een complete herindexering van het GGC waarbij ook nieuwe velden en nieuwe sleutels worden toegevoegd wordt ingepland voor 2005.

Dat de indexen nu in Unicode kunnen worden gemaakt, betekent niet dat dat ook echt gebeurt. Per index (auteursindex, titelindex, titelsleutelindex etc.) kan nl. ingesteld worden of de indexen gebaseerd zijn op Unicode of op de wijze zoals in Pica3 gebruikelijk was. Zo is het bijvoorbeeld nu mogelijk om aparte indexen te maken op "élevé" en "élève" en op "the" en "thé". Het is echter niet per se nodig om dat te doen.

Bij het zoeken zonder accenten worden in principe alle index-ingangen gevonden. Dus zoeken met "eleve" levert zowel "éleve", als ook "élevé" en ook "élève". Zodra met een acent gezocht wordt, wordt ook letterlijk met die string gezocht. Dus zoeken met "élevé" levert niet "élève".

Maar de exacte inductabellen zullen nog kunnen veranderen en zullen in november op basis van de tests nog definitief worden vastgesteld. Hierover worden de GGC-gebruikers apart geïnformeerd.

2.3 Unicode en LBS3

De titeldatabase in LBS3 is gebaseerd op de Pica-characterset en niet op Unicode.

De record-uitvoer ten behoeve van het update-mechanisme voor LBS3 (OUM) is gebaseerd op de Pica-characterset. In het OUM zal een Unicode-switch worden aangebracht. Zolang het update—mechanisme niet expliciet vraagt om Unicode-records zullen de records worden uitgevoerd in de Pica-characterset – en zullen de records dus worden teruggeconverteerd. Zodra het OUM niet meer om Pica-characterset vraagt, zullen de records in Unicode worden aangeleverd.

Komen er inmiddels in de records in het GGC Unicode-tekens voor die niet in de Pica-characterset ondersteund worden en wordt door het OUM gevraagd om uitvoer in de Pica-characterset, dan zullen de Unicode-tekens worden verwijderd uit de records, en worden deze dus niet aangeleverd aan het LBS.

Over de bouw en uitlevering van deze nieuwe OUM-versie zullen de gebruikers nog apart worden geïnformeerd.

2.4 Unicode en OUF

Gebruikers die voor de online export van records uit het GGC gebruik maken van het OUF (Online Update Fetch, de gestandaardiseerde versie van het OUM) bevinden zich in een soortgelijke situatie als de gebruikers die ten behoeve van het LBS gebruik maken van OUM. Ook in OUF wordt de mogelijkheid ingebouwd om de gewenste characterset van de records nader te bepalen. In de huidige OUF-versie is de uitwisseling van records gebaseerd op de Pica-characterset en verandert de export derhalve niet.

Ook hier geldt dat indien er in de records in het GGC Unicode-tekens voorkomen die niet in de Pica-characterset ondersteund worden en indien gevraagd wordt om uitvoer in Pica-characterset, dat dan de Unicode-tekens zullen worden verwijderd uit de records, en dus niet worden aangeleverd.

Over de implementatie van de nieuwe versie van OUF zal met de betreffende gebruikers bilateraal contact worden opgenomen.

2.5 Unicode en uitvoer

De offline uitvoer bestaat in twee varianten: XML-uitvoer en lijstuitvoer. De XML-uitvoer is reeds in Unicode en zal dus niet wijzigen. De lijstuitvoer gebruikt de Pica-characterset en dat zal ook zo blijven. Bij lijstuitvoer zullen Unicode tekens worden teruggeconverteerd naar de Pica-characterset en Unicode tekens die niet in de Pica characterset voorkomen worden weggefilterd.

3 Set sortering

Zoals hierboven gemeld is de nieuwe zoek- en indexengine in het CBS ingebouwd. Afgezien van de Unicode-faciliteiten beschikt deze nieuwe software ook over de mogelijkheid om sets te sorteren. Vooralsnog zullen de standaard-sorteerfaciliteiten die ook in PiCarta worden aangeboden in GGC en NCC/IBL worden aangeboden. Dat biedt de mogelijkheid om sets alfabetisch te sorteren op titel of auteur. Deze faciliteit werkt uitsluitend bij sets kleiner dan 1000 hits.

Voor auteurs werkt het sorteercommando – involgorde van aanwezigheid - op kmc 3000, 300X en 3100. Voor titels werkt het sorteercommando op kmc 4000 (Hoofdtitel).

Er kan een plus- of een minteken worden toegevoegd om oplopend of aflopend te sorteren.

Voor de uitbreiding van de sorteerfaciliteiten, bijvoorbeeld bij ingangsrecords, zal een apart plan worden gemaakt voor invoering in 2005.

Sets sorteren werkt op basis van het SORT-commando.

Voorbeeld:

- sor s3 aut
- sor ttl
- sor s6 aut –

4 Zoeken met woordcombinaties

De nieuwe indexmachine in het CBS maakt het ook mogelijk om bij het zoeken met titelwoorden de volgorde van woorden mee te laten wegen. Met het zoekcommando: z ttl "american journal" worden alle records gevonden waarin deze woorden naast elkaar voorkomen. Met het zoekcommando: z ttl "american journal of archeology" worden alle records gevonden waarin deze woorden voorkomen. En met het zoekcommando: z ttl "journal of archeology" worden zowel de Amerikaanse als Europese versie gevonden.

5 IBL

5.1 Velden in bibliotheekadressen 1 regelig

Tot aan deze versie kunnen in onderdelen van gebruiker- en bibliotheekadressen sommige velden langer zijn dan 1 regel. Zo kunnen meerdere e-mail-adressen worden ingevoerd gescheiden door "nieuwe regel". In de software die gebruik maakt van deze velden worden deze meervoudige regels echter niet opgevangen en niet goed afgehandeld. Vandaar dat besloten is om met ingang van versie 2.1 deze velden te beperken tot één regel. Voorzover bekend heeft deze ingreep geen nadelige gevolgen voor de huidige NCC-gebruikers.

5.2 Bibliotheektype "onbekend"

Sedert de invoering van de nieuwe tariefsdifferentiatie in de vorige CBS-versie wordt het veld "bibliotheektype" gebruikt voor vaststelling van de IBL-tarieven. Bij inspectie is gebleken dat niet bij alle bibliotheek- en gebruiker-records dit veld is ingevuld. Dit leverde problemen op bij de presentatie en afhandeling van dit gegeven. Derhalve is in het systeem een nieuw bibliotheektype "onbekend" geïntroduceerd voor het geval er geen bibliotheektype is ingevuld. Voor de vaststelling van de IBL-tarieven zal in dergelijke gevallen het standaard-IBL-tarief gebruikt worden.

5.3 Bonnen en antwoordcodes via e-mail

In het IBL-systeem is een faciliteit aanwezig om IBL-aanvragen te versturen via de zgn. IBL-gateway. In versie CBS 2.1 wordt deze faciliteit verder uitgebouwd. Zo is een nieuw formaat ingevoerd (Pica ILL) dat nu geheel gelijk is aan het formaat dat op de zgn. "compacte bon" gebruikt wordt. Bovendien is de faciliteit voor "status-and-error-reports" uitgebreid met status-informatie. Een gedetailleerde beschrijving van deze functionaliteit is beschikbaar en verkrijgbaar via de Servicedesk.

5.4 Korte presentatie deposito's

In de korte presentatie van gebruiker- en bibliotheekdeposito's kwamen de tekens ">", "<" en "!" voor, voor respectievelijk "actief", "niet actief" en "geblokkeerd". Deze tekens zijn uit de presentatie verwijderd in verband met fouten bij het opbouwen van schermpresentaties. De status als zodanig is niet uit de deposito's verwijderd. Om de status te raadplegen kan nog steeds het volledige deposito-record geraadpleegd worden.

5.5 Onderlinge verrekening

De rapporten voor de Onderlinge Verrekening zijn aangepast en zijn nu gebaseerd op de bedragen die in de IBL-aanvragen worden opgeslagen. Het opslaan van die financiële gegevens in de IBL-aanvragen gebeurt al sedert 1 juli 2003.

Deze nieuwe software voor de Onderlinge Verrekening maakt deel uit van versie 2.1, maar gezien het uitstel van deze versie tot november is besloten om dit deel van de software eerder te implementeren en reeds in november te gebruiken voor het opstellen van de rapporten voor de Onderlinge Verrekening.

Een tweede aanpassing met betrekking tot de Onderlinge Verrekening betreft een fout die optrad bij het opslaan van positief afgehandelde IBL-aanvragen. In dat geval wordt de vlag "Onderlinge verrekening J/N" niet goed weggeschreven. Deze fout doet zich voor bij aanvragen van bibliotheken die een deposito hebben bij een andere bibliotheek. De fout is hersteld in versie 2.1 en zal voor de aanvragen waarbij dit verkeerd is gegaan met een script worden hersteld. Voor de aanvragen die onderdeel zijn van het derde kwartaal van de Onderlinge Verrekening is dit inmiddels al gebeurd.

5.6 Selecteren van IBL-aanvragen

In het IBL-systeem kunnen IBL-aanvragen worden beantwoord met antwoordcodes 1 t/m 9. IBL-aanvragen kunnen sinds versie 2.0 nadat ze bijvoorbeeld positief zijn gehonoreerd ook nog de status RETURNED, CHECKED-IN en RECEIVED ontvangen. Deze codes of statussen maken onderdeel uit van het standaard ISO IBL-protocol. Met ingang van versie 2.1 kan ook op deze statussen gezocht worden.

6 Formaatuitbreidingen

Het Pica3 formaat is uitgebreid met een aantal nieuwe kmc's en subvelden. Verder zijn er enkele andere aanpassingen in het formaat aangebracht.

Deze kmc's worden opgesomd in de tabel hieronder.

Een beschrijving van de kmc's en de aanpassing van de Richtlijnen voor deze kmc's zijn al verspreid via de Listserver GGC.

Omschrijving kmc	Pica3 kmc	Pica+ kmc
Taal	1500	010@
ISBN	2000, 2001, 2005, 2009	004A, 004B, 004C, 004D
Verwant ISSN	2011	005C
UPC nummer	2067	006Q
Actuele uitgever	4029	033M
Link tussen elektronische en papieren versie van tijdschrift	4280	039T
Exemplaargebonden uitgever	4830	233A
GOO-periode	533X (X = 1-9)	045T/0X (X = 1-9)

7 Statistiek

7.1 TOT in plaats van t/m

Bij het in WinIBW opvragen van statistiek van instellingen over een periode langer dan één maand (toon stat feb 2003 tot apr 2003) werd "tot" geïnterpreteerd als "tot en met". Dit wordt in CBS-versie 2.1 gecorrigeerd.

7.2 Elektronische aanvragen

In het CBS waren tot nu toe aparte tellers gereserveerd voor het indienen van elektronische aanvragen. In praktijk worden deze tellers gebruikt om het doorklikken naar full text te registreren in de eindgebruikerbestanden. Aangezien dit doorklikken in het CBS (GGC en NCC/IBL) niet mogelijk is, worden deze tellers dan ook met CBS versie 2.1 uit het CBS verwijderd.

7.3 Tellers 201-212

In januari 2003 is een fout in de statistiektellers geslopen bij de tellers 201-212. Dit betreft de tellers die bij het catalogiseren het ontlenen aan derden (BNB, LC, etc.) registreren. Onlangs is ontdekt dat deze tellers in 2003 en 2004 verkeerd zijn opgehoogd. In plaats van de broncode (= instellingsnummer) van het bibliografisch record werd de broncode van de catalogiserende user gecontroleerd. Deze fout wordt in CBS versie 2.1 gecorrigeerd.

7.4 Teller bij inloggen

De teller die wordt gebruikt bij het inloggen was tot op heden niet beschikbaar in het datawarehouse dat gebruikt wordt voor de webstatistiek. Dat is aangepast zodat ook de resultaten van deze teller zichtbaar kunnen worden gemaakt.