

## **Høring i danZIG af:**

### ***Udlægning til og ordning af registre i onlinekataloger***

På behandlingen på møde i Bibliografisk Råd af forslag til *Udlægning til og ordning af registre i onlinekataloger* blev der rejst et forslag om at anvende standard UNICODE mekanismer til sortering samt rejst spørgsmålet om forholdet mellem søgning og sortering.

Da medlemmerne af BIR ikke umiddelbart kunne tage stilling til om anvendelse af DUCET (Default Unicode Collation Element Table) ville være hensigtsmæssigt, blev beslutningen i BIR, at det tekniske spørgsmål sendes til udtalelse i danZIG samt at den nedsatte arbejdsgruppe vurderer kommentarerne fra danZIG og kommer med en fornyet indstilling til BIR.

Sagen kan kort resumeres således:

a) På mødet i BIR refererede en af DBCs repræsentanter til et forslag fra Tommy Schomacker om, at der anvendes DUCET (Default Unicode Collation Element Table) ved at der vælges eller laves en national variant, som bl.a. sikrer at æ, ø og å sorteres korrekt. Der henvises til *Default Unicode Collation Element Table # (DUCET) for the Unicode Collation Algorithm*

<http://www.unicode.org/Public/UCA/latest/allkeys.txt> .

På mødet i BIR havde Statsbibliotekets repræsentant en henvisning til ICU - International Components for Unicode <http://site.icu-project.org/home>

b) Søgning af ord og sætninger skal uden diskussion ske på den udlagte normaliserede form, som beskrevet i udkastet til specifikation.

c) Sortering af ordregistre og sætningsregistre kan så ske på følgende to måder:

- Sortering efter den udlagte normaliserede form
- Sortering i forhold til den ikke-normaliserede form med brug af DUCET

Dette spørgsmål ønskes drøftet i danZIG med henblik på efterfølgende behandling i arbejdsgruppen bag forslag til *Udlægning til og ordning af registre i onlinekataloger*.

d) Sortering af symboler (afsnit 3.4) er et særligt spørgsmål, hvor de forskellige "standardkomponenter" angiveligt giver forskellig rækkefølge. Dvs. vi skal under alle omstændigheder give en regel for hvorledes symbolerne skal sorteres, så det bliver ens i "alle" danske onlinekataloger. Arbejdsgruppen har ingen gode argumenter til brug for at favorisere den ene standardkomponent frem for den anden. Måske lige bortset fra at vi ikke bør vælge en, der er afhængig af et enkelt programmeringssprog.

I forslaget til *Udlægning til og ordning af registre i onlinekataloger* angives denne rækkefølge:

\$%&\*+<=>@ç£¥§©¬®°±·×÷∞

Det blev nævnt i forbindelse med BIR-mødet at med ICU's collator (med dansk locale), fås denne rækkefølge:

·@\*&#%§©®°+±÷×<=>¬∞×ç\$£¥

Hvis der i stedet benyttes Javas indbyggede collator fås:

§©®@×ç\$£¥\*&#%+±÷×<=>¬°∞

e) Alfnummerisk sortering i vis-formater er udenfor forslag til *Udlægning til og ordning af registre i onlinekataloger*. Tommys forslag indebærer at den nationale variant af DUCET anvendes til dette formål. Ønskes drøftet i danZIG.

f) I forlængelse af ovenstående ønskes drøftet om anvendelse af DUCET (Default Unicode Collation Element Table) i givet fald skal angives som obligatorisk eller om det alene skal anvises som et relevant værktøj.

Tommy vil præsentere et oplæg på danZIG-mødet, hvor han bl.a. vil beskrive hvordan DUCET kan tilrettes til formålet.

PS:

Modtagne høringssvar fra danZIG-medlemmer:

Flemming:

AXIELL Danmark A/S har modtaget forslaget til "Udlægning til og ordning af registre i onlinekataloger" til høring i danZIG.

Vi har ingen kommentarer til forslaget.

Vi vil gøre opmærksom på at de 2 systemer, som AXIELL sælger på det danske marked (DDElibra og AXIELL Arena), pt. ikke er i stand til at følge de 2 væsentligste ændringer i forslaget: supplerende normalisering (§4.4) og dublerende udlægning af titler mv. indledt med ordene nævnt i §9.1.2. Dette gælder formodentlig også andre systemer på markedet.

Per:

Kan ikke lige bidrage med noget i denne forbindelse. Så altså – ok fra mig.

Tommy:

Blot for en ordens skyld. DBC har naturligvis ingen principielle kommentarer. Eventuelle redaktionelle småting vil blive samlet op i BIR. Personligt mener jeg, at der bør skelnes mellem søgning og sortering. Søgning handler om udlægning af flere søgestrengene for det samme (Aalborg og Ålborg). For sortering bør vi overveje om vi kan anvende standardmekanismerne i unicode.