

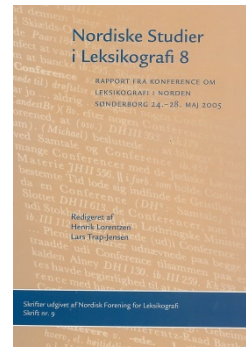
# NORDISKE STUDIER I LEKSIKOGRAFI

**Titel:** Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy

**Forfatter:** Oddrun Grønvik og Lars Jørgen Tvedt

**Kilde:** Nordiska Studier i Leksikografi 8, 2006, s. 143-150  
Rapport från Konferens om lexicografi i Norden, Göteborg 27.-29. maj 1999

**URL:** <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

## Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

## Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

## Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy

Norsk Ordbok 2014 – presentation of a complex lexicographic tool. The dictionary writing (DWS) system of Norsk Ordbok has been developed by the project Norsk Ordbok 2014 and the Unit for digital documentation (EDD) at the University of Oslo. This paper briefly describes the main modules of this DWS.

### 1. Innleiing

Norsk Ordbok skal gjerast ferdig innan 2014. For å nå dette målet vart prosjektet Norsk Ordbok 2014 (NO 2014) oppretta, med ein utvida redaksjonsstab. Det vart naudsynt å innføre nye metodar og nye verktøy for å effektivisere redigeringsprosessen.

Målsettinga med arbeidet vart også endra. I tillegg til å vere ei papirordbok, er det no eit mål at Norsk Ordbok skal gjerast tilgjengeleg i digital form.

Med dette som utgangspunkt har prosjektet NO 2014 og Eining for digital dokumentasjon (EDD) ved Universitetet i Oslo utvikla eit spesiallaga redigeringsssystem for ordboka.

Redigeringsystemet har vakse fram i eit tett samarbeid mellom NO 2014 og EDD. Spesifiseringsarbeid og utviklingsarbeid har gått parallelt, og nye modular har blitt utvikla og innarbeida i systemet samstundes som redigeringsystemet har vore i bruk i redaksjonen.

Vi vil i denne artikkelen kort skissere hovudfunksjonane til dette nye redigeringsystemet. Dette er ikkje ein fullstendig dokumentasjon eller presentasjon av systemet, men ei kortfatta innføring i kva hovudmodular vi har vald å ta med i systemet.

### 2. Generelt system

EDD utviklar system for ei rad ulike fagområde, og i avtalen mellom NO 2014 og EDD vart det lagt som premiss at dei systema som vart utvikla i samarbeid mellom NO 2014 og EDD skulle kunne brukast også av andre. EDD valde difor å basere redigeringsystemet på det rammeverket som alt var utvikla. Dette avsnittet skildrar difor dette rammeverket i generelt.

#### 2.1. *Metadatabase*

Etter som talet på ulike fagdatabasar vaks innanfor ramma av EDD, vart det stilt større og større krav til gjenbruk av komponentar for å halde kostnadane ved systemutviklinga lågast mogleg. Det vart også stilt krav om at ein burde kunne søkje

i ulike databasar utan å måtte bruke ulike databaseapplikasjonar. EDD fann det difor naudsynt å vurdere om ein kunne utvikle eit system som sentraliserte alle felles delar av eit større databasesystem, men samstundes gjorde det mogleg å utvikle det spesielle for kvar einskild fagdatabase som EDD fekk ansvaret for.

Løysinga på dette problemet var å identifisere alle objekt i databasen med ein objekttype og ein identifikator. Med objekt meiner vi til dømes ein ordboksartikkel, ein setel i ein seteldatabase, ein kataloginnføring frå eit museum, eller ei lydfil frå ein musikkdatabase.

EDD utvikla så ein felles metadatabase med informasjon om strukturen for dei ulike fagdatabasane, eit felles Windows-rammeverk med tilsvarende web-rammeverk, og eit system for å kople spesialutvikla framvisarar/redigeringsskjema til dei ulike objekttypane.

Metadatabasen som ligg til grunn for systemet, er basert på ein relasjonell databasestruktur. For kvar ny database ein utviklar, registrerer ein metadata om tabellar og relasjonar slik at ein har meir informasjon om dette enn det som ligg i den normale metadata-strukturen for eit databasesystem. Denne informasjonen brukar systema dynamisk til å generere søkjeskjema, og søka mot databasen vert genererte på basis av metadata og data frå brukaren.

Resultata av slike spørjingar vert presenterte som resultatlistar som får si form via data i metadatabasen. I tillegg vert resultata vist i spesialkonstruerte redigeringsskjema. Koplinga mellom resultatlistar og redigeringsskjema er også styrt via metadatabasen.

Spørjesystemet representerer i utgangspunktet relasjonsmodellen som ligg til grunn for dei ulike databasane. Alle datafelt i alle tabellar er i utgangspunktet tilgjengeleg som søkjefelt for sluttbrukaren. For dei fleste brukarar av systemet vil dette framstå som komplekst og utilgjengeleg. EDD har difor utvikla eit system i metadatabasen som gjer det mogleg å forenkle strukturen slik at den passar for kvar einskild. Dei fleste databasane er difor utstyrde med ein eller fleire spesiellaga søkjeskjema som brukaren kan velje mellom. I tillegg er det fritt for brukaren å lage sine egne søkjeskjema. Dette er mogleg fordi ein har basert seg på dynamisk bruk av data frå metadatabasen i staden for å generere statisk programkode eller skjemastruktur, slik mange tilsvarende system gjer.

Den einskilde brukaren kan også lagre søkjekriteria eller resultat av søk i det ein har kalla lister. Dette gjer det mogleg å kome lett tilbake til dei objekta i basen som ein arbeider med, eller som har spesielle karakteristikkar. Dette har vore mykje brukt i statushandtering i samband med registrering, innskriving, osv.

Sidan databasane til EDD varierer i stor grad, fann ein ut at det ikkje ville vere naturleg å lage felles redigeringsskjema for alle basane, eller forsøke å lage dynamisk generering av redigeringssystema. I staden har EDD utvikla eit generelt grensesnitt mellom det dynamiske søkjesystemet og spesiellaga redigeringsskjema for dei ulike

databasane. Søkjesystemet bruker metadatabasen til å kontrollere kva redigerings-skjema som høyrer til kvar database.

Ved å kombinere dynamisk generering av søkjesystema med spesiellaga redigerings-skjema, har EDD utvikla ein struktur som gjer det mogleg å fokusere på det spesielle ved kva einskild base, samstundes som det meste er generelt. EDD kan difor utvikle avanserte system på kort tid.

Basert på same metadatabase-struktur har EDD også utvikla eit web-system med same funksjonalitet som redigeringsapplikasjonen har. Søkjeskjema og resultat er lett gjenkjenneleg for brukaren, men i staden for redigerings-skjema, får ein her fram ei HTML-vising av objekta ein leiter fram.

## **2.2. Teknologi**

I botnen av fellessystemet ligg ein Oracle-database. Oracle er vald både ut frå funksjonsomsyn, men også fordi Universitetet i Oslo har ein avtale med denne databaseleverandøren.

På klientsida har EDD så langt berre utvikla løysingar for operativsystemet Microsoft Windows og for web. Grunnen til dette er at universiteta i Noreg stort sett brukar Windows. EDD har nytta Delphi frå Borland som utviklingsplattform.

Sidan EDD må handtere meir enn berre standardtekst, har ein tatt i bruk diverse ulike teknologiar. I databasen lagrar EDD både strukturert tekst, lyd, bilete og kartdata, og alle desse datatypene vert manipulert både i databasesystemet og på klienten.

## **3. Redigeringsystemet**

Redigeringsystemet til Norsk Ordbok er utvikla innanfor det rammeverket som er omtalt i førre avsnitt. Strukturen for å bygge opp søkjesystem og kopling til andre databasar låg difor klar før modelleringa av ordboksstrukturen tok til.

### **3.1. Grunnlagsmateriale**

Mykje av grunnlagsmaterialet for Norsk Ordbok er digitalisert og ligg som eigne fagdatabasar i det felles rammeverket som er utvikla ved EDD. Norsk Ordbok 2014 har brukt ein god del ressursar på å normalisere dette materialet ved å kople det saman via Metaordboka som også er ein fagdatabase i rammeverket. Dette er mogleg fordi objekta er unikt identifisert via objekttypar og identifikatorar.

### **3.2. Struktur**

Norsk Ordbok er modellert strengt hierarkisk. Dette vart gjort fordi ein i utgangspunktet rekna med at dette er den rette strukturen for ordboka. Etter kvart har ein vel kome fram til at det finst strukturar som bryt med denne føresetnaden, og dette har skapt visse problem.

Under føresetnad av ein hierarkisk struktur på artiklane i ordboka, er redigerings-systemet for artiklane basert på ein trestruktur som speglar databasestrukturen. Redigering av dei einskilde delane av artikkelen vert gjord i skjema knytt til den einskilde noden. Mykje av kontrollen med at redigeringsreglane vert følgde har vi lagt inn i programvaren i databasetenaren. Dette har vi gjort for å kunne sikre dataintegritet sjølv om ein nyttar ulike verktøy for å manipulere data i databasen.

Ved å dra nytte av at alle objekt er identifisert via objekttypen og ein identifikator, kan ein i teorien kople alle typar data til aktuelle nodar i artikkelreet. Dette gjer det mogleg å dokumentere alle tydingar, eller andre delar av artikkelen, ved å peike til andre data, som til dømes setelarkiv, andre ordbøker, osv.

### **3.3. Presentasjon**

Det er lagt vinn på at så lite som mogleg av skiljeteikn, bindetekstar, og liknande, skal leggjast inn i databasen. Presentasjon av artikkelen i form av trykte boksider eller i form av webpublisering, skjer ved at data frå databasen vert eksportert til eit format som er tilpassa presentasjonsformatet. Vi har basert oss på at artiklane vert eksportert som XML. For generering av websider vert XML transformert ved hjelp av XSLT til HTML, og mest mogleg av formatering av HTML vert gjort ved hjelp av CSS.

For papirpresentasjon, har vi basert oss på å bruke PDF. Artiklane vert konvertert frå XML til XSL-FO ved hjelp av XSLT, og dette vert så formatert ved hjelp av XSL-FO-programvare til PDF. Vi har forsøkt å bruke gratisprogramvare til dette, mellom anna FOP, men på grunn av spesielle krav som NO 2014 stilte, har vi enda opp med å bruke XEP frå RenderX for transformasjon til PDF. Heile trykksatsen for band 5 vart generert på denne måten som ei PDF-fil.

Eit av problema i samband med papirversjonen av ordboka, var handtering av spesialteikn. Norsk Ordbok valde å nytta fonten Lexicon frå The Enschedé Font Foundry, og Enschedé konstruerte dei manglande teikna i denne fonten.

Eit anna stort problem var handtering av orddeling. I ei ordbok med mange avstyttingar er det ikkje mogleg å lage rutinar for korleis dette skal gjerast. I tillegg til dei regelbaserte systema kom vi difor ikkje unna store ordlister med unntaksreglar. Vi måtte også lage reglar for å hindre linjedeling mellom ord. Alt dette vart lagt inn i XSL-FO-systema våre.

### **3.4. Redigeringshjelp**

#### *3.4.1. Autogenerering av artikkelskjelett*

Redigeringsystemet til Norsk Ordbok drar nytte av arbeidet som er gjort med Metaordboka ved at ein automatisk genererer opp eit rammeverk for ein artikkel basert på innhaldet i Metaordboka. Systemet tolkar kjeldematerialet som er kopl til metaordboksartiklane, og gjev redaktørane råd om dei bør lage artiklar i Norsk

Ordbok basert på dei einskilde metaordboksartiklane. Vel ein å lage ein ny artikkel i Norsk Ordbok, lagar systemet ein artikkel med oppslagsformar og ordklassemerking på plass. I tillegg vert peikarar til grunnlagsmaterialet kopla til artikkelen, slik at store delar av dokumentasjonsarbeidet er gjort alt før redaktøren har skrive ei einaste tyding.

#### *3.4.2. Sortering og kopling av grunnlagsdata*

I redigeringsystemet til Norsk Ordbok er det lagt inn eit sorteringsverktøy som gjer det mogleg å sortere grunnlagsmaterialet etter dei kriteria som til ei kvar tid er nyttig for den aktuelle artikkelen. Sorteringsinformasjonen vert teken vare på så lenge redaktøren sjølv ønskjer, og basert på sorteringsinformasjonen kan redaktøren bygge opp tydingsstruktur med kopling direkte til dei einskilde kjeldene. I tillegg kan ein brukar av ordboka sjå kva setlar frå setelarkivet, eller kva korpusbelegg som ligg til grunn for ein kvar definisjon i ordboka.

Redaktøren kan også hente inn anna grunnlagsmateriale undervegs i redigeringa av artikkelen. På dette viset kan redaktøren mellom anna kople korpusbelegg direkte til tydingar i artikkelen.

#### *3.4.3. Krysstilvisingar*

Ein sentral funksjonalitet i eit redigeringsystem er å halde kontroll på alle tilvisingar i systemet. Redaktørane kan berre vise til artiklar eller tydingar som reelt eksisterer i ordboka. Dette er sikra ved at alle tilvisingar er reelle koplingar til artiklar som er redigert i databasen. Det er ikkje mogleg å slette ein artikkel som ein annan refererer til. Den hierarkiske visinga av artikkelen syner tilvisingar i begge retningar, slik at det mellom anna er mogleg å sjå kva artiklar som viser til den artikkelen ein redigerer.

#### *3.4.4. Handtering av kjelder og heimfestingar*

For å sikre konsistens i handteringa av heimfesting av tydingar, døme og belegg, er det lagt inn eit stadhierarki i systemet som automatisk skjuler heimfestingar som er omfatta av andre heimfestingar for same tyding. Redaktøren kan også be systemet om å sjekke om det finst så mange ulike heimfestingar at det bør slåast saman til ei heimfesting på høgare nivå. Til dømes kan systemet automatisk slå saman ei rad kommunar til eit fylke.

Det er også meininga at systemet etter kvart skal kunne lese ut alle heimfestingar og kjelder frå grunnlagsmaterialet, og på dette viset spare redaktørane for arbeidet med å redigere dette sjølv.

#### *3.4.5. Dimensjonering*

Norsk Ordbok er etter kvart komen i grenselandet mellom å vere ei digital ordbok

og ei prenta ordbok. I den digitale verda er ikkje volum eit stort problem, men den trykte utgåva må halde seg innanfor visse rammer. I redigeringsystemet er det difor lagt inn eit system for automatisk å foreslå storleiken ein artikkel skal ha. Utrekninga av dette er basert på innhaldet i Metaordboka for den tilsvarende artikkelen. Talet på belegg for ein metaordboksartikkel vert lagt til grunn, og basert på diverse vektingsreglar vert storleiken for artikkelen i Norsk Ordbok rekna ut. Det talet som kjem fram er ei rettesnor for redaktørane, men dette talet kan sjølvsagt overprøvast. Undervegs vert den reelle storleiken på artikkelen registrert, og dette kan danne grunnlag for rapportar.

#### *3.4.6. Førehandsvising*

Trass i at dei fleste redaktørane etter kvart har funne seg til rette med å sjå på artiklar som hierarkiske tre, er det sjølvsagt mogleg for redaktøren å sjå korleis artikkelen vil sjå ut i papirform eller i web-form. Redaktøren kan velje å kontinuerleg vise ein HTML-basert versjon av artikkelen, eller artikkeldelar, og kan også generere ut PDF-versjon av ein eller fleire artiklar.

### *3.5. Administrasjon*

For eit prosjekt med over tjue redaktørar, er det naudsynt med administrasjon. Vi har lagt vinn på å legge så mykje som mogleg av administrasjonen inn i systemet på ein slik måte at administreringa vert ein del av dei daglege oppgåvene til redaktørane.

Redaksjonen i Norsk Ordbok er inndelt i grupper, med ein eller fleire gruppeleiarar. I tillegg er det oppretta nokre spesialgrupper innanfor redaksjonen. Desse spesialgruppene har særskilte oppgåver som til dømes å sjekke sitat eller samanhalde større delar av ordboka. I systemet er dette løyst ved at redaktørane er delt inn i redaksjonsgrupper, og i tillegg har roller.

Alle artiklar har til ein kvar tid ein redaktør. I tillegg til dette har alle artiklar ein status. Statussystemet er laga slik at ein artikkel med ein bestemt status berre kan flyttast over i eit avgrensa tal andre statusverdiar. Statusen til artikkelen avgjer kven som har tilgang til å redigere han, og kven som kan flytte artikkelen mellom ulike statusar. Ved hjelp av dette systemet sikrar ein at ikkje fleire redaktørar redigerer same artikkel. Samstundes kan ein bruke systemet til å bygge opp ei produksjonsline der artikkelen må innom ein rad stasjonar før han vert publisert. Til dømes må ein artikkel innom ein gruppeleiar for kontroll, og alle artiklar må ha vore innom sitatkontroll før hovudredaktør tar tak i totalen.

### *3.6. Logging*

Det er eit mål at redaksjonen til ein kvar tid har oversikt over kva som har skjedd med ein artikkel. Det er difor konstruert eit loggsystem knytt til statushandteringa. Kvar gong ein artikkel endrar status, vert det teken ein kopi av artikkelen med

statusinformasjon, og denne kopien vert liggande i databasen. Det er også mogleg å ta ein slik kopi kvar gong redaktøren ønskjer det. Dette gjer det mogleg å følgje utviklinga av artikkelen, både innhaldsmessig og også for å undersøkje kven som har gjort kva med artikkelen.

### 3.7. *Statistikk*

Basert på statusverdiar, dimensjoneringstal og loggar, vert det no manuelt laga produksjonsrapportar for at administrasjonen skal kunne følgje utviklinga av ordboka. Dette er viktig for å halde ein høg produksjon. Vi arbeidar no med eit generelt rapporteringssystem som skal gjere det lettare å utvikle dei rapportane som redaksjonen treng.

## 4. Problem

Ein del av problema vi har støtt på i utviklinga av systemet, er å strukturere ordboka, samstundes som ein skal opne for fleksibilitet. Ved å bruke eit relasjonsdatabasesystem i botnen, har vi lagt ganske sterke strukturingskrav på ordboka.

Vi har også hatt debattar om korleis redigeringa skal gå føre seg. I ein tidlegare versjon av systemet var trestrukturen meir gjennomgripande ved at det hovudsakleg berre var mogleg å redigere einskildnodar. I dagens versjon er det laga ein kombinasjon av ein trestruktur og eit vanleg skjema som eigentleg omsluttar fleire noder i treet. Den hybridløysinga som no er vald skapar ein del problem, og etter som redaksjonen vert meir van med å tenkje artikkelen som eit tre, vil det kanskje vere ønskjeleg å gå over til å kunne redigere nodane i treet meir direkte. I dette ligg det i så fall eit krav om å flytte fokus endå meir bort frå satsbiletet og meir i retning av strukturering av innhaldet.

## 5. Konklusjon

Redigeringsverktøyet slik det no ligg føre er eit godt verktøy til å fullføre det store NO 2014-prosjektet, sjølv om delar av systemet ikkje fungerer heilt slik vi kunne ønskje. Vidare – sjølv om det framleis står att å utvikle nokre modular av systemet – ser vi likevel at systemet har ført til høgare kvalitet, meir konsistens i ordboka, større effektivitet og enklare opplæring av nye redaktørar.

## Litteratur

Extensible Stylesheet Language (XSL), Version 1.0, W3C Recommendation 15 October 2001, <http://www.w3.org/TR/xsl/>  
 XSL Transformations (XSLT) Version 2.0, W3C Working Draft 15 September 2005, <http://www.w3.org/TR/xslt20/>



**Web-sider:** (alle sider sjekka 4. oktober 2005)

Cascading Style Sheets Home Page (CSS), <http://www.w3.org/Style/CSS/>

Eining for digital dokumentasjon, <http://www.edd.uio.no/>

Formatting Objects Processor (FOP), <http://xmlgraphics.apache.org/fop/>

Norsk Ordbok, <http://no2014.uio.no/>

RenderX, <http://www.renderx.com/>

The Enschedé Font Foundry, <http://www.teff.nl/>

The Extensible Stylesheet Language Family (XSL), <http://www.w3.org/Style/XSL/>