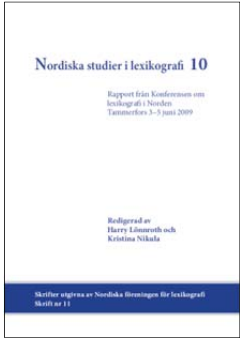


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Forbedring af datatilgang i elektroniske opslagsværker i forbindelse med kognitive brugersituationer	
Forfatter:	Marcin Overgaard Ptaszynski	
Kilde:	Nordiska Studier i Lexikografi 10, 2010, s. 417-431 Rapport från Konferens om lexicografi i Norden, Tammerfors 3.-5. juni 2009	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

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 lexicografi (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.

MARCIN OVERGAARD PTASZYNSKI

Forbedring af datatilgang i elektroniske opslagsværker i forbindelse med kognitive brugersituationer

It would not be incorrect to claim that besides developing theories and methodologies that can improve the compilation of needs adapted reference works, the ultimate objective of lexicography is to offer the user of a reference work quick and easy access to data which satisfy his need for information. Therefore, access to data should reflect the user's needs which appear in specific extra-lexicographical situations. This article offers an investigation of access to data in electronic reference works which aim to satisfy the needs that appear in the so-called cognitive situations (cf. Tarp 2006: 57 ff.). The analysis focuses on forms of access that hitherto rarely have been discussed by lexicographers. On its basis it is concluded that access to data can be and should be improved. This requires both a redefinition of the term *access* and a more efficient implementation of available technological solutions.

Nøgleord: tilgang, søgevej, søgestreng, søgeresultat, elektronisk leksikografi, kognitiv situation, brugerbehov, leksikografisk funktion

1. Indledning

Hvor mange oktaver kan et klaver spille? Hvis man ikke kender det rigtige svar, men gerne vil kende det, kan man henvende sig til en pianist eller besøge en klaverforhandler. Alternativt kan man vælge at søge svaret i et opslagsværk. I så fald har man et leksikografisk relevant behov. Dette behov kan betegnes som et behov for at tilegne sig viden om et bestemt emne, og ifølge den moderne leksikografiske funktionsteori opstår den i en kognitiv situation. Behovet kan tilfredsstilles ved hjælp af et opslagsværk, der indeholder relevante data, for eksempel en encyklopædi eller en musikordbog.

I det følgende vil jeg undersøge tilgang til data i elektroniske leksikografiske produkter der søger at tilfredsstille brugerens behov som opstår i bestemte typer kognitive situationer. For at undersøgelsen kan gennemføres, bliver det dog først nødvendigt at kigge nærmere på begrebet tilgang og foreslå en ny definition af den. Tilgang er nemlig mere nuanceret end der bliver gjort udtryk for i den eksisterende litteratur om emnet. Ligeledes må forskellige typer kognitive situationer undersøges grundigere.

På baggrund af analysen vil der blive argumenteret for at tilgang til data kan og bør tilpasses bedre til brugerens behov. Da den nødvendige teknologi dertil findes allerede, består forbedring af tilgangsformerne først og fremmest i at udnytte den optimalt. Der bliver derfor også kigget på i hvor høj grad det allerede er sket i udvalgte nordiske opslagsværker.

2. Tilgang til data i elektroniske leksikografiske produkter

2.1 *Tilgang og brugerbehov – en metaleksikografisk skitse*

Begrebet *tilgang* (*til data*) og tilsvarende begreber på andre sprog (eng. *access* og ty. *Zugriff*, *Datenakzessivität*) bruges temmelig ofte i (meta)leksikografi (se f.eks. Wiegand 1988, 2008, Hausmann & Wiegand 1989, De Schryver 2003, Bergenholtz & Gouws 2007). Opfattelsen af tilgang har for det meste været kendetegnet ved en strukturalistisk tankegang, hvilket kan ses i de følgende definitioner:

Die Zugriffsstruktur (eines Nachschlagewerkes) Z_s ist eine (endliche, nichtleere) Menge, die Menge der Leitelemente M_{Le} , zusammen mit einer Zugriffsrelation R_o auf M_{Le} , also $Z_s = \langle M_{Le}; R_o \rangle$. (Wiegand 1988: 56.)

Tilgangsstruktur er en struktur af de leksikografiske indikatorer, der leder brugeren frem til de søgte oplysninger. (Nielsen 1993: 240.)

The access structure can be regarded as the search route the dictionary user follows during a dictionary consultation procedure. (Steyn 2004: 276.)

Tilgang defineres således gennem søgeveje og strukturer i en ordbog, som brugeren benytter sig af for at søge efter data. Denne brug af termen *tilgang* synes imidlertid ikke at være dækkende. For det første passer den først og fremmest til opslagsværker i papirformat. I elektroniske ordbøger og encyklopædier, især de internetbaserede, er mediostrukturen og "Leitelemente" ikke nødvendigvis relevante for brugeren, der kan få adgang direkte til de søgte data ad en anden søgevej end gennem lemmalisten. For det andet er det at betragte tilgang som en søgevej eller en struktur lige med at reducere dens opfattelse til leksikografens perspektiv. For brugeren er det derimod ikke bare vigtigt at kende søgevejen til de relevante data. Lige så relevante er for eksempel spørgsmålene om hvordan man bør formulere sin søgning, hvilke resultater man får ved at benytte sig af

konkrete søgningsmuligheder, eller hvor meget tid det tager at nå frem til de oplysninger man efterlyser. Det peger på at begrebet tilgang dækker over en proces, hvor strukturen kun er et element. Henning Bergenholtz og Rufus Gouws (2007: 240) synes at dele denne mening, idet de taler om ”the notion of an access process [...] which includes the access structure as one of its components”. Definitionen af tilgang som proces trænger dog til at blive præciseret. Som et forsøg på dette vil jeg opstille følgende definition:

Tilgang er en proces, hvorved brugeren af et leksikografisk produkt målrettet søger og finder data, der umiddelbart tilfredsstillende hans konkrete leksikografisk relevante behov. Processen er brugerinitieret, brugermotiveret og brugerstyret, og den foregår i samspil med en række af opslagsværkets indre og ydre strukturer. Den begynder ved brugerens udvælgelse af opslagsværket og formulering af søgningen, den foregår via en konkret søgevej, og den slutter når brugeren når frem til de eftersøgte data der findes blandt de søgeresultater som foreligger som svar til søgningen.

Adskillige aspekter af tilgang i elektroniske opslagsværker er blevet drøftet i leksikografisk litteratur, og interessen for emnet er berettiget. I sin artikel der giver et overblik over udvikling i elektronisk leksikografi AD 2003, påpeger Gilles-Maurice De Schryver følgende:

The true core of the ED [electronic dictionary – MOP] revolution [...] lies in the fact that users are liberated from the alphabetical straitjacket, that hypertext, menus, etc. eliminate (artificial) linear text restrictions, that the data conjured up on screen are not static, and that powerful search capabilities ensure a smooth overarching navigation. (De Schryver 2003: 157.)

samt:

It is the variety of look-ups possible in EDs (fuzzy searches, hyperlinks, etc.) which makes this medium truly exciting. (De Schryver 2003: 155.)

Tilgang som proces er dog ikke endnu blevet etableret som et forskningsemne i sig selv. Den nævnes ofte i andre sammenhænge (her begrænset til elektronisk leksikografi), først og fremmest i anmeldelser af elektroniske ordbøger, men også i beskrivelser af udfærdigede eller igangværende leksikografiske projekter, og i studier der også omfatter andre aspekter af leksikografi. Påfaldende er det dog at tilgang til data sjældent diskuteres i forbindelse med konkrete bru-

gerbehov. Den nærværende artikel er således et forsøg på både at vie tilgang den forskningsinteresse den fortjener og at relatere den til nogle bestemte typer brugerbehov, nemlig dem, der opstår i sporadiske kognitive situationer. Jeg vil begynde med at rette fokus mod søgevejtyper.

2.2 Søgeveje i elektroniske opslagsværker

Hvordan søges der i elektroniske opslagsværker? Der findes fire grundlæggende typer af søgeveje, afhængigt af deres startpunkt, det vil sige det sted i eller udenfor produktet, hvor brugeren begynder sin søgning. For det første kan man foretage en søgning gennem lemmalisten, hvilket kan gøres i de fleste ordbøger på cd- eller dvd-rom og nogle internetbaserede ordbøger og leksika, for eksempel *Aschehougs Leksikon* (2009). For det andet kan man søge data gennem en fortegnelse over begrebskategorier. Denne type søgevej findes i opslagsværker, hvis makrostruktur er tematisk opbygget, så som de fleste nyhedsportaler på internettet, eller delvist tematisk opbygget, for eksempel *Wikipedia* (2009). For det tredje kan man få adgang til data gennem hyperlinks placeret i opslagsværkets artikler og omtækster, eller helt udenfor ordbogen (f.eks. på en ikke-relateret hjemmeside eller i andet software). Hyperlinkene kan være skjulte, men de kan også være gjort synlige for brugeren, for eksempel ved hjælp af understreg eller brug af farver. De kan lede brugeren til data placeret i opslagsværkets artikler, omtækster, eller helt udenfor opslagsværket. For det fjerde og sidste kan man foretage en søgning gennem et søgefelt. Dette kan være universalt, det vil sige maskinen foretager altid søgning i alle datatyper uden at brugeren har muligheden for at justere søgningen (f.eks. i *Gyldendals Røde Ordbøger* 2009). Et søgefelt kan dog også være definerbart, hvilket betyder at man kan justere sin søgning med hensyn til forskellige parametre. Man kan for eksempel begrænse den til konkrete byggede af opslagsværket eller konkrete datatyper, man kan udforme søgestrengen på en bestemt måde (f.eks. bruge booleske operatoren), eller man kan definere sin søgning i forhold til den type leksikografisk relevant behov man har. Mange leksikografiske produkter tilbyder brugeren flere typer søgeveje. I *Wikipedia* (2009) finder man for eksempel en lemmaliste, en fortegnelse over begrebskategorier og et definerbart søgefelt.

På trods af forskellene imellem dem, har de ovennævnte søgevejtyper meget til fælles, især hvis man betragter dem fra en metaleksikografisk synsvinkel. Først og fremmest er søgevejene som regel beskrevet i de produkter, hvor de tilbydes. En undtagelse her er søgningen via universalt søgefelt i de produkter der ikke tilbyder andre søgeveje. Derudover har alle søgevejtyperne været gen-

stand for videnskabelig analyse, oftest i forbindelse med anmeldelser af konkrete opslagsværker der tilbyder dem. Sidst men ikke mindst kendetegnes næsten alle de omtalte typer søgeveje ved at de blev udarbejdet ud fra andre parametre end konkrete typer brugerbehov for information. Den eneste undtagelse er et søgefelt hvor brugeren ikke alene skal udforme en konkret søgestreng, men også tilpasse sin søgning til den konkrete type ekstra-leksikografisk situation han befinder sig i.¹ Denne type søgevej sikrer en nemmere og hurtigere tilgang til relevante data og forefindes for eksempel i *Base lexicale du français* (2009) eller i internetordbøger udarbejdet efter principper rodfæstet i den moderne leksikografiske funktionsteori (se bl.a. Bergenholtz & Tarp 2003, Tarp 2006), for eksempel *Ordbøgerne over Faste Vendinger* (2009) og *Musikordbogen* (2009).

3. Søgeveje og kognitive situationer

De hidtil omtalte typer søgeveje i elektroniske opslagsværker er ikke de eneste der forefindes. De udtømmer heller ikke det videnskabelige potentiale der er blevet leveret af leksikografiske teorier, herunder funktionsteorien. I *Musikordbogen* (2009) er behov der stammer fra kommunikative situationer (tekstreception) ganske vist blevet adskilt fra behov der affødes af kognitive situationer. I *Ordbøgerne over Faste Vendinger* (2009) finder man endda søgemuligheder opdelt efter ordbogsfunktioner der svarer til enkelte typer kommunikative situationer. De kognitive situationer afspejles derimod i én, samlet ordbogsfunktion: at hjælpe brugeren med at tilegne sig mere viden om et bestemt emne. Indtil videre findes der ingen leksikografiske produkter hvor der bevidst og konsekvent bliver taget højde for de forskellige slags kognitive situationer. Denne tilstand skyldes flere faktorer, men for at forstå problemet til bunds er det først og fremmest nødvendigt for leksikografen at anerkende at der findes flere forskellige kognitive situationer.

Sven Tarp (2006: 59) fordeler kognitive situationer i to kategorier: sporadiske og systematiske. De førstnævnte kendetegnes ved at den leksikografisk relevante problemstilling er punktuel, det vil sige det drejer sig om et behov for information der kan tilfredsstilles ved hjælp af en præcis, nøjagtigt afgrænset mængde information. For eksempel kan man under en diskussion om politik

1 I den nærværende artikel henvises der til brugeren som en abstrakt størrelse, og ikke som en konkret person. For at opnå en konsekvent fremstilling i teksten bruges der derfor kun ét genus af de relevante personlige og possessive pronomener, og det er altid maskulinum.

få et have behov for at vide hvornår Polen blev medlem af EU. Svaret er enkelt: 1. maj 2004. I systematiske kognitive situationer er problemstillingen derimod globalt, og behovet for information kan ikke tilfredsstilles med et enkelt, konkret stykke data tilgængeligt et bestemt sted i opslagsværket. Ofte kan den ikke tilfredsstilles ved en enkelt opslagsaktivitet. Et eksempel på sådan en situation kan være et behov for at vide hvordan Europaparlamentet fungerer, der opstår i forbindelse med forberedelse til en eksamen i Europas historie.

Tarp (2008: 124) bemærker at opdelingen af kognitive situationer i sporadiske og systematiske har konsekvenser for opbygningen af en ordbog. Hvad der ikke bliver nævnt, er hvordan man lige præcis kan tilbyde brugeren den mest effektive form for tilgang i forhold til den konkrete type situation. Endvidere kan det siges at Tarps opdeling af kognitive situationer ikke er helt overbevisende hvis den skulle fungere som en egentlig taksonomi. Der er snarere tale om en gruppering af eksempler der skal illustrere de to overordnede typer kognitive situationer. Opdelingen kan alligevel have en betydning for tilgangsproces. Brugeren der befinder sig i sporadiske kognitive situationer har nemlig et behov for at vide mere, men ikke nødvendigvis alt, om et bestemt emne. Hvordan kan man tilbyde ham en nem tilgang til netop de data han har brug for? Problemet gør sig gældende især i de artikler der beskriver centrale begreber eller kendte personer, for eksempel *opera* eller *Chopin* indenfor musik, eller i artikler om lande, for eksempel *Polen* eller *Danmark*. I flere leksikografiske produkter vil man se at disse artikler indeholder en betragtelig datamængde. Det er dog langt fra altid tilfældet at brugeren har behov for at tilgå alle disse data. Hvis man blot vil vide hvad årsagen til Chopins død var, er oplysninger om hans barndom, ophold i Frankrig, ja sågar al hans musik, irrelevante for én.

4. Alternative former for tilgang

Der er tre faktorer som i væsentlig grad medvirker til at der ikke findes nogen tilfredsstillende leksikografisk løsning på problemet med behovstilpasset tilgang til data i sporadiske kognitive situationer. For det første findes der former for tilgang som ikke er tilstrækkelige, så som anvendelsen af mini-indholdsfortegnelser i artikler. For det andet findes der former for tilgang, for eksempel fritekstsøgning i HTML tekst, som kun sjældent fanger metaleksikografernes interesse, og som sjældent bliver udnyttet af leksikografer. For det tredje gør man ikke tilstrækkelig brug af den nye teknologi der er nødvendig for at tilbyde nye, bedre former for tilgang. Da hver af de tre faktorer rummer et forbedringspotentiale, fortjener de en mere detaljeret beskrivelse.

4.1 Mini-indholdsfortegnelser

En måde hvorpå man kan hjælpe brugeren med at finde hurtigt frem til de ønskede data er gennem mini-indholdsfortegnelser placeret i artiklen. Denne type data og dens anvendelse i papirordbøger beskrives af for eksempel B. T. Sue Atkins og Michael Rundell (2008: 204). Et eksempel på et elektronisk leksikografisk produkt som har en kognitiv funktion og som også gør brug af denne løsning er *Wikipedia* (2009). Her består indholdsfortegnelsen af overskrifter, der fungerer som hyperlinks, det vil sige brugeren bliver omdirigeret til det pågældende afsnit i artiklen ved at klikke på overskriften.

Umiddelbart synes anvendelsen af indholdsfortegnelsen med henblik på at få hjælp i sporadiske kognitive situationer at have to fordele. For det første bliver den tid det tager brugeren at nå til de ønskede data reduceret. Brugeren behøver for eksempel ikke at gennemlæse hele artiklen om Wagner for at finde frem til hvorfor hans musik blev foretrukket af nazisterne i Tyskland. Et relevant hyperlink forekommer nemlig i indholdsfortegnelsen. For det andet er søgevejen integreret i produktet. Man kan oven i købet deaktivere den hvis man skal bruge produktet i forbindelse med systematiske kognitive situationer. Hvis brugeren som udgangspunkt er interesseret i at tilegne sig mere omfattende viden om Wagner og dermed læse hele artiklen, er den interaktive indholdsfortegnelse nemlig unødvendig.

På den anden side er denne søgevej ofte langt fra tilstrækkelig til at sikre nem og hurtig adgang til data. I nogle artikler er indholdsfortegnelsen så udvidet at det alligevel tager tid at finde frem til netop den overskrift man har brug for (se f.eks. den komplekse fortegnelse i artiklen *opera* i den engelske udgave af *Wikipedia* 2009). Derudover er det ikke givet at de data man efterlyser, er placeret netop under den overskrift hvor man formoder at finde dem (og dette kan skyldes både en dårlig artikelopbygning og brugerens utilstrækkelige evner til at relatere sin søgen til en bestemt overskrift i indholdsfortegnelsen). Sidst men ikke mindst kan det ske at brugeren vælger den relevante overskrift, men det pågældende afsnit indeholder langt flere data om emnet end man lige har brug for.

Indholdsfortegnelsen, hvis den skal tilbydes, må holdes indenfor overskuelige grænser. Disse er svære at fastslå, men man kan med rimelighed sige at en fortegnelse inddelt i tre eller flere niveauer kan tage meget tid at gennemlæse, hvilket muligvis vil få brugeren til at foretrække en anden type søgevej. På den anden side kan begrænsning af indholdsfortegnelsens længde og kompleksitet have den stik modsatte effekt især i virkelig lange artikler. Jo færre overskrifter i indholdsfortegnelsen, jo større mængden af data placeret under hver overskrift i artiklen.

4.2 Fritekstsøgning i HTML tekst

I internetbaserede opslagsværker findes der faktisk allerede en relativ enkel måde at reducere den tid der er nødvendig for at komme frem til de ønskede data. Den eksisterer dog helt uden om de bestræbelser som leksikografen har gjort sig for at gøre produktet brugervenligt, og dermed også uden om metaleksikografernes opmærksomhed. Måden består i at bruge en specifik tastekombination for fritekstsøgning i HTML tekst (som regel Ctrl+F). Den anvendes almindeligvis til at finde frem til alle forekomster af en konkret søgestreng på en internetside.

Fritekstsøgning kan med fordel bruges i længere ordbogsartikler, hvis brugeren hurtigt vil nå frem til de data han søger i en sporadisk kognitiv situation. For at illustrere denne form for tilgang, lad os forestille os en bruger der vil finde ud hvor mange oktaver et klaver kan spille. Behovet er leksikografisk relevant og burde kunne tilfredsstilles af et passende produkt, for eksempel *Musikordbogen* (2009). Efter at have indtastet *klaver* i ordbogens søgefelt og valgt muligheden ”jeg vil gerne vide mere” (der svarer til kognitive situationer generelt) bliver brugeren præsenteret for den lange artikeltekst. I stedet for at skulle læse definitionen af *klaver* og oplysninger om hvordan et klaver virker, kan brugeren foretage fritekstsøgning i artiklen og søge direkte efter søgestrengen *oktav*. Internetbrowseren finder den første forekomst af søgestrengen og fremhæver dens baggrund. Søgestrengen viser sig faktisk at være præcis den, som brugeren leder efter, idet den optræder i sætningen ”Et moderne klaver har som regel et omfang på 7 oktaver [...]”.

Fordelen ved benyttelse af fritekstsøgning som søgevej til de efterlyste data er at den giver brugeren en mulighed for at reducere søgetiden betydeligt. Ulempen er at mens man foretager søgningen, forbliver hele artikelteksten synlig, hvilket kan virke distraherende. Ved længere artikler risikerer man nemlig at miste overblikket over teksten og dermed søgeprocessen. Derudover er søgevejen ikke integreret i produktet da det kun eksisterer i internetbrowseren. Sidst, men ikke mindst er fritekstsøgning ikke altid helt pålidelig. Browseren finder ganske vist alle forekomster af søgestrengen som optræder i artikelteksten, men også alle de der ellers findes i det samme aktive browservindue. Dermed får en bruger der søger efter *EU* i artiklen *Polen* på *Wikipedia* (2009) også sådanne forekomster som for eksempel ”(D)eu(tsch)” eller ”Eu(skara)” blandt søgeresultaterne. Og det på trods af at disse forekomster vises i et andet område af browservinduet end det, hvor man finder artikelteksten. Dette problem kan dog løses hvis fritekstsøgning bliver integreret i produktet.

4.3 ”Udnyttet teknologi”

Ud over at forbedre indholdsfortegnelser i artiklerne eller integrere fritekstsøgning i produktet kan leksikografen forbedre tilgang til data på en tredje måde. Med dagens teknologi kan man nemlig udarbejde en form for tilgang der er mere raffineret end den der gør brug af de hidtil omtalte søgevejtyper. Inspiration kan hentes både fra eksisterende opslagsværker og fra internetsøgemaskiner, for eksempel Google. Man kan nemlig tilbyde en søgevej i form af et integreret søgefelt der tillader boolesk søgning (og det findes i en lang række elektroniske ordbøger, for eksempel *Nationalencyklopedins Ordbok* 1997, samt encyklopædier, for eksempel *Wikipedia* 2009) og samtidig præsentere søgeresultater som en række synopsis fra en eller flere artikler, sorteret efter relevans.

For at illustrere ideen, lad os forestille os nogen der under en diskussion vil afgøre hvor mange strenge der kan være i en guitar. Her vil det være oplagt at kunne benytte et søgefelt, hvor det er muligt at søge udelukkende efter de relevante artikler i encyklopædien, det vil sige dem der indeholder ordene, eller rettere sagt, bogstavkombinationerne *guitar* og *streng*. Som søgeresultatet kunne man få en liste synopsis fra alle de relevante artikler, og hver synopse ville indeholde netop de(t) ord der udgør søgestrengen. Denne løsning er i sig selv ikke noget nyt. Den anvendes i høj grad i *Wikipedia* (2009), men den fungerer ikke altid efter hensigten.

Alternativt kunne det være muligt at søge direkte og udelukkende efter de uddrag i artiklen *guitar* som indeholder bogstavkombinationen *streng*. Søgeresultatet kunne i så fald bestå af en liste af de synopsis fra artiklen *guitar* hvor *streng* optræder, og synopsiserne ville hver især for eksempel bestå af den sætning der indeholder kombinationen *streng*. Denne form for tilgang kunne især være fordelagtig for en bruger der ved (eller har en fornemmelse af) hvor i opslagsværket han vil søge svar på sit spørgsmål. Endvidere, hvis søgefeltet integreres i produktet således at der er en direkte adgang til den fra artiklerne (og ikke kun fra forsiden), vil denne form for tilgang også være ideel for tilfredsstillelse af andre sporadiske kognitive behov, for eksempel dem der opstår under konsultationen af opslagsværket hvis man pludselig ønsker at få mere viden om det pågældende emne (jf. Tarp 2006: 59).

Der er flere fordele ved denne form for datatilgang. Den reducerer både tid og besvær forbundet med søgning efter de ønskede data. Brugeren behøver nemlig ikke at gennemlæse hele artiklen, men kun vælge det uddrag der indeholder de data han søger. Søgevejen er også integreret i produktet, hvilket gør det muligt at undersøge den form for tilgang ved hjælp af analyse af logfiler.

Ulemper ved en sådan form for tilgang ligger på den tekniske side af produktudviklingen. At tilbyde sådanne synopsis fra artikelteksten der varierer

afhængigt af søgestrengen ville kræve en database der kan programmeres, for eksempel MySQL. Den ville også kræve mere arbejde fra forfatterne til artiklerne. Med udgangspunkt i brugerens potentielle behov for konkret information må artikelteksten nemlig opdeles i mindre dele og hver del må oprettes som et datafelt.

5. Hvor langt er vi kommet?

For at se hvor godt tilgang til data i nutidens elektroniske opslagsværker er tilpasset brugerbehov der opstår i sporadiske kognitive situationer har jeg valgt at kigge nærmere på fem nordiske internetbaserede encyklopædier: *Store Norske Leksikon* (2009), *Den Store Danske* (2009), danske *Aschehougs Leksikon* (2009), svenske *Nationalencyklopedin* (2009), og den danske udgave af *Wikipedia* (2009). Med udgangspunkt i den løsning der blev foreslået i sidste afsnit har jeg stillet 12 lukkede spørgsmål med hensyn til hvert af de ovennævnte værker. Spørgsmålene vedrører følgende aspekter af tilgang: søgevejen, søgestrengen og søgeresultatet. Både spørgsmålene og svarene præsenteres i tabellen i bilaget. Her vil jeg blot beskrive proceduren i undersøgelsen og nævne generelle konklusioner som kan drages fra den.

I hvert af de undersøgte opslagsværker har jeg forsøgt på at foretage boolesk søgning i det i produktet integrerede søgefelt. Søgestrengen bestod altid af to bogstavkombinationer, adskilt med mellemrum og en boolesk operator (medmindre bruget af denne var frarådet i brugervejledningen), for eksempel ”oktav + klaver”. Søgestrengen kunne variere fra det ene opslagsværk til det andet, således at jeg kun søgte efter ”Hamsun + Hitler” i *Store Norske Leksikon* (2009), og ikke for eksempel i *Aschehougs Leksikon* (2009), hvor strengen var ”oktav klaver” (ingen operatører skal benyttes her). Denne strategi blev anvendt dels med henblik på at sikre sig at de relevante data fandtes i det pågældende opslagsværk (f.eks. var Knut Hamsun nordmand, og derfor kunne man forvente at en norsk encyklopædi har flere oplysninger om ham end man kan finde i en dansk eller svensk encyklopædi), dels ud fra den betragtning at formålet var at undersøge *tilgangen* til data, og ikke dataenes tilstedeværelse i et bestemt opslagsværk. Hver søgestreng afspejlede dog en reel ekstra-leksikografisk sporadisk kognitiv situation, for eksempel ”oktav + klaver” var udtrykket for et behov for at vide hvor mange oktaver et klaver kan spille.

Dernæst har jeg undersøgt søgeresultatet af hver søgning med henblik på at afgøre hvorvidt den er tilpasset det sporadiske kognitive behov der lå til grund for søgningen. Jeg har især været interesseret i om søgeresultatet bestod af synopsis fra relevante artikler, og om synopsiserne var udformet således at de var

relevante for søgningen, det vil sige om de indeholdt de(t) ord der udgjorde søgestrengen.

Undersøgelsens resultater lader en drage to generelle konklusioner. Den ene er positiv, idet både søgevejene og søgeresultaterne i alle de undersøgte opslagsværker har potentiale for at blive tilpasset brugerbehov der opstår i sporadiske kognitive situationer. Alle fem encyklopædier tillader nemlig boolesk søgning, og i alle på nær *Aschehougs Leksikon* (2009) består søgeresultatet af en liste synopser fra relevante artikler. Samtidig viser undersøgelsen dog at såvel de søgeveje der tilbydes som udformningen af søgeresultaterne kan forbedres betydeligt.

Skønt boolesk søgning er tilladt i alle opslagsværkerne, virker den ikke altid i overensstemmelse med brugervejledningen. I nogle encyklopædier, for eksempel *Wikipedia* (2009), skal man bruge andre operatoren end dem som brugervejledningen anbefaler. I andre opslagsværker, for eksempel i *Store Norske Leksikon* (2009), skal man helt undgå at bruge operatoren, selv om brugervejledningen siger det modsatte. Endvidere viser det sig at rækkefølgen af de bogstavkombinationer der udgør søgestrengen kan påvirke søgeresultatet, så at søgningen ”oktav (+) klaver” og ”klaver (+) oktav” giver forskellige resultater.

Selv om fire ud af de fem undersøgte opslagsværker tilbyder søgeresultater i form af en liste over stikord efterfulgt af synopser, er synopserne ikke altid relevante i forhold til søgestrengen. Dette kan skyldes at synopserne er faste i forhold til specifikke artikler, hvilket er tilfældet med for eksempel *Store Norske Leksikon* (2009). Selv hvis synopserne er variable, betyder det ikke at denne løsning anvendes konsekvent (se f.eks. *Wikipedia* 2009). Variationen kan også være meget begrænset i sit omfang og eksistere stik imod de oplysninger man finder i brugervejledningen (se f.eks. *Nationalencyklopedin* 2009). Sidst, men ikke mindst er det ikke givet at søgeresultatet består af synopser hentet fra alle artikler der er relevante for søgningen.

De ovennævnte u hensigtsmæssigheder bør naturligvis udbedres. Den komparative undersøgelse af opslagsværkerne viser klart at der allerede findes den nødvendige teknologi. Optimering af tilgang til data i forbindelse med sporadiske kognitive situationer kan da foretages relativt nemt.

6. Slutbemærkning

Kernen i denne artikel er beskrivelsen af tre forskellige former for tilgang til data i elektroniske opslagsværker: brug af indholdsfortegnelser i artikler, fritekstsøgning i HTML tekster, og anvendelse af boolesk søgning til at få variable, relevante synopser som søgeresultater. Disse former for tilgang er på sin

vis alternative, idet de enten ikke endnu har tiltrukket megen opmærksomhed fra leksikografer, eller de gør brug af teknologiske løsninger der er tilgængelige men som ikke udnyttes optimalt. De tre former for tilgang rummer samtidig et potentiale for forbedring.

På det teoretiske plan bør det siges at leksikografernes interesse for tilgang ikke må være begrænset til udarbejdelsen af optimale strukturer for placering og lokalisering af sprogligt og sagligt korrekte data. Den bør også omfatte andre aspekter af tilgang, så som udformning af søgestrengene, udarbejdelse af nye søgeveje og behovstilpasset præsentation af søgeresultater. Tilgang skal forstås som en proces, og ikke bare som en struktur, og dens former skal tilpasses til bestemte brugerbehov.

LITTERATUR

Ordbøger og andre opslagsværker

- Aschehougs Leksikon, 2009. <http://www.aschehougsleksikon.dk/>
 Base lexicale du français, 2009 = Lexical Database for French (Base lexicale du français – BLF). <https://ilt.kuleuven.be/blf/>
 Den Store Danske, 2009 = Den Store Danske. Gyldendals åbne encyklopædi. <http://www.denstoredanske.dk/>
 Gyldendals Røde Ordbøger, 2009 = Gyldendals Røde Ordbøger. <http://ordbog.gyldendal.dk/>
 Musikordbogen, 2009 = Inger Bergenholtz: Musikordbogen. Århus. <http://www.musikordbogen.dk/>
 Nationalencyklopedin, 2009 = Nationalencyklopedin. <http://www.ne.se/>
 Nationalencyklopedins ordbok, 1997 = Nationalencyklopedins ordbok på cd-rom. Malmö.
 Ordbøgerne over Faste Vendinger, 2009 = Bergenholtz, Henning under medvirken af Esben Bjærge og Richard Almind: Ordbøgerne over Faste Vendinger. Århus. <http://www.idiomordbogen.dk/>
 Store Norske Leksikon, 2009 = Store Norske Leksikon. <http://www.sn.no/>
 Wikipedia, 2009 = Wikipedia. The Free Encyclopedia. <http://www.wikipedia.org/>

Anden litteratur

- Atkins, Beryl T. S. & Rundell, Michael, 2008: The Oxford Guide to Practical Lexicography. Oxford.
 Bergenholtz, Henning & Gouws, Rufus, 2007: The Access Process in Dictionaries for Fixed Expressions. I: *Lexicographica* 23. S. 237–260.

- Bergenholtz, Henning & Tarp, Sven, 2003: Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. I: *Hermes* 31. S. 171–196.
- De Schryver, Gilles-Maurice, 2003: Lexicographers' Dreams in the Electronic-dictionary Age. I: *International Journal of Lexicography* 16: 2. S. 143–199.
- Hausmann, Franz Josef & Wiegand, Herbert Ernst, 1989: Component Parts and Structures of General Monolingual Dictionaries: A Survey. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales handbuch zur Lexikographie*, red. af Franz Josef Hausmann et al. Berlin/New York. S. 328–360.
- Nielsen, Sandro, 1993: Tilgangsstruktur. I: *Manual i fagleksikografi*, red. af Henning Bergenholtz & Sven Tarp. Herning. S. 240–245.
- Stejn, Mariza, 2004: The Access Structure in Learner's Dictionaries. I: *Lexikos* 14. S. 275–298.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. Doktorafhandling ved Handelshøjskolen i Århus.
- Tarp, Sven, 2008: The Third Leg of Two-legged Lexicography. I: *Hermes* 40. S. 117–141.
- Wiegand, Herbert Ernst, 1988: Wörterbuchartikel als Text. I: *Das Wörterbuch. Artikel und Verweisstrukturen*, red. af G. Harras. Düsseldorf. S. 30–120.
- Wiegand, Herbert Ernst, 2008: Zugriffstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. I: *Lexicographica* 24. S. 209–316.

Bilag. Testspørgsmål og testresultater i undersøgelsen af tilgang til data i udvalgte encyklopædier

	Store Norske Leksikon (2009)	Aschehougs Leksikon (2009)	Den Store Danske (2009)	National- encyklopedin (2009)	Wikipedia (DK) (2009)
1. Findes der et søgefelt?	+	+	+	+	+
2. Er boolesk søgning tilladt?	+	+	+	+	+
3. Forklares anvendelsen af søgevejene i en bruger-vejledning?	+	+	-	+	+
4. Virker boolesk søgning i overensstemmelse med instruktionen i bruger-vejledningen?	- ¹	+	N/A	- ²	- ²
5. Påvirkes søgeresultatet af rækkefølgen af søgeordene i søgestrengen?	-	+ ³	+ ⁴	-	-
6. Bliver der søgt efter afgrænsede eller uafgrænsede bogstavkombinationer?	afgr.	uafgr.	afgr.	afgr.	(uafgr.) ⁵
7. Viser der flere søgeresultater (hvis der forekommer flere)?	+	+	+	+	+
8. Viser listen over søgeresultaterne i centrum af skærmen?	+	-	+	+	+
9. Viser der synopsis i søgeresultatet?	+	-	+	+	+
10. Er synopsiserne altid relevante (dvs. indeholder de søgte ord)?	-	N/A	-	-	-
11. Varierer synopsiserne afhængigt af søgestrengen?	-	N/A	-	(+) ⁶	(+) ⁵
12. Er de søgte ord fremhævet i søgeresultaterne?	-	+ ⁷	+ ⁷	+ ⁸	+ ⁸

afgr. = afgrænset (bogstavkombination)
uafgr. = uafgrænset (bogstavkombination)

¹ Søgning med de anbefalede operatorer giver ingen resultat.

² De anbefalede operatorer opfattes som en del af søgestrengen.

³ Både antal af de fundne relevante artikler, og synopsisernes gensidige placering i søgeresultatet er påvirket.

⁴ Søgningen påvirkes også af brug af booleske symboler. Både antal af de fundne relevante artikler, og synopsisernes gensidige placering i søgeresultatet er påvirket. Således giver søgning efter ”klaver + oktav”, ”oktav + klaver”, ”klaver oktav” og ”oktav klaver” fire forskellige søgeresultater.

⁵ Løsningen virker ikke konsekvent.

⁶ Variationen har et meget begrænset omfang, og eksisterer stik imod beskrivelsen præsenteret i brugervejledningen.

⁷ I artiklerne, men ikke i synopsiserne.

⁸ I synopsiserne, men ikke i artiklerne.