

En hale uten ende

Nasjonalbibliotekets pilotprosjekt om nordområdene – en evaluering etter ett år
12. juni 2008



"... older content tends to score higher because it's had longer to accumulate incoming links. In other words, search inverts the usual priority of content; older is often the better."

[Chris Anderson – "The Long Tail", 2006]

1. Bakgrunn for pilotprosjektet

Norske forfattere, forlag, tidsskriftutgivere og andre rettighetshavere og Nasjonalbiblioteket har som felles mål å gjøre litteratur og annet rettighetsbelagt materiale tilgjengelig, samt gi allmennheten innblikk i og kunnskap om mangfoldet i den norske kulturarven.

Som et ledd i nå dette målet, har Nasjonalbiblioteket og representanter for norske rettighetshavere¹ ett års tid samarbeidet om et pilotprosjekt der norske verk gjøres digitalt tilgjengelig over internett innenfor et nærmere avtalt emneområde – **nordområdene**. Målsettingen med pilotprosjektet er blant annet å høste erfaringer med teknologiske løsninger, lenkeproblematikk, avtalemekanismer, brukeradferd m.m., erfaringstilfang som kan danne grunnlaget for eventuelle framtidige avtaler om digital formidling av rettighetsbelagt materiale.

Vilkårene for samarbeidet er nedfelt i et eget avtaledokument, se <http://www.nb.no/nordomradene/>.

Pilotprosjektet ble lansert 24. april 2007. Etter ett år har deltakerne ønsket å evaluere hvordan brukerne vurderer tilbudet. Denne korte rapporten oppsummerer resultatene fra en kvantitativ analyse av brukeradferd per 24. april 2008. En mer utførlig evalueringsrapport vil foreligge etter pilotperiodens utløp i oktober/november.

2. Beskrivelse av funksjonalitet

Definisjoner

I denne prosjektrapporten defineres følgende begreper slik:

Med **bruker** menes en person med unik IP-adresse².

Et **besøk** skjer når en bruker åpner et digitalt dokument for nærmere studium.

En **visning** av en eller flere boksider eller tidsskriftsider pågår når en bruker "besøker" ett verk og studerer siden/-ene på dataterminal. En side "vises".

¹ Rettighetshaverne er representert ved Den norske Forleggerforening, Den norske Forfatterforening, Norsk faglitterær forfatter- og oversetterforening, Norsk kritikerlag og Forvaltningsorganisasjonen LINO. LINO har etter særskilt avtale med BONO, GRAFILL, Forbundet Frie Fotografer og Norges Fotografforbund fått fullmakt til å inngå avtale på vegne av medlemmene i disse organisasjonene i pilotprosjektet.

² Bak hver IP-adresse kan det skjule seg flere personer, fordi en bedrift, organisasjon eller lignende kan operere med kun én IP-adresse.

Søk og gjenfinning i NBdigital

Hovedinngangen til innholdet i NBdigital er gjennom søk i Nasjonalbibliotekets søkemotor. Fra søkevinduet kan brukeren lete i over 3 millioner oppføringer: nettsider, katalogposter og mer enn 1 million digitale objekter.

Nasjonalbibliotekets søk er et integrert søk, som gjør oppslag i mange forskjellige kildedatabaser³, på tvers av materialtype og faggrenser. Brukeren søker i hele samlingen samtidig, og trefflistene som fremkommer, sammenstiller dermed innhold og opplysninger på en ny måte. Samtidig genereres det dynamiske navigatorer basert på metadataene som følger objektene, slik at brukeren kan forfine treffet ved hjelp av strukturerte data.

Fulltekstvisning av bøker og tidsskrifter

Bøkene i NBdigital OCR- og strukturbehandles, og påføres supplerende metadata. Teksten som fremkommer ved OCR-behandlingen, indekseres i NBs søkemotor sammen med metadata.

Det er altså mulig å søke på tradisjonelt vis etter tittel og forfatter (metadata), og på samme tid søke etter ord og fraser i selve innholdet (fritekst). Ved søketreff i teksten kommer brukeren inn på de sidene i boka der frasen ble funnet og kan bla videre derfra. Det gjøres også en automatisert strukturanalyse der en eventuell innholdsfortegnelse annoteres, og der sidenummer i boka verifiseres slik at den digitale visninga forholder seg til den faktiske pagineringa.

Brukeren kan bla i verket, gå til et gitt sidenummer og velge mellom forskjellig kvalitet og visningsmåter. Sidene som vises fram er jpg-bilder av reelle boksider. De opptrer som enkeltbilder. Det er ikke lagt til rette for nedlasting av hele bøker eller OCR-tekst.

Bibsys-brukere kan gå direkte fra bokvisning til bestilling av boka gjennom Bibsys. Tidsskrifter scannes i sin helhet, men det legges begrensninger på tilgangen slik at brukeren bare kan lese artikler som er rettighetsklarerte innenfor avtaler.

Nordområdematerialet som egen samling

Nordområdematerialet får ei særskilt merking i NBs katalog, slik at objektene kan opptre som ei egen samling i det digitale biblioteket. Det er mulig å avgrense søket til bare å gjelde denne delmengden av indeksen.

Nordområdeportal

Når et brukersøk gir treff i materiale som er merket "nordområdene" i katalogen, eller søkeordet finnes på ei predefinert liste med nordområderelaterte emneord, aktiveres et "supertreff" som er inngangen til portalsida om Nordområdene: <http://www.nb.no/nordomradene/> Her settes Nordområdebegrepet i en breiere kontekst, med aktuelle nyheter, artikler, presentasjoner og ei stor lenkesamling til aktuelle nordområderessurser i hele verden.

Synlighet

Nordområdematerialet er eksponert mot eksterne søkemotorer slik at robotene finner og indekserer innholdet. Dette gjør materialet søkbart utenfor NBdigital.

Bortsett fra redaksjonell omtale i dagspressen, etermedia og bransjetidsskrift er ikke tjenesten bekjentgjort via avisannonsering, TV-reklame eller annen betalt markedsføring.

³ Per 1. april 2008 består den digitale samlingen av:

200 000 aviser / 365 000 bilder / 50 000 bøker / 500 filmer / 1000 timer musikk / 300 000 timer radio / 250 tidsskrifter

3. Erfaringer etter ett års drift

Brukerundersøkelsen er basert på 394 boktitler og 230 tidsskriftartikler (til sammen 634 titler).

Antall unike sider: 81468 (71425 boksider, 10048 artikkelsider).

Antall unike IP-adresser: 2482

332 av brukerne (13,4%) har brukt både bøker og tidsskrift.

Antall besøk: 3798 (bøker) og 1463 (artikler).

Antall verk besøkt: 308 (bøker) og 192 (artikler), henholdsvis 78% og 83%

Samlet antall sider vist: 23014 (bøker) og 4739 (artikler)

Antall unike sider vist: 10535 (bøker) og 1622 (artikler)

Antall sider vist per besøk (snitt): 5,23 (bøker) og 2,02 (artikler)

I 56% av titlene er 1 - 20% av innholdet vist fram.

Bruk etter utgivelsesår - basert på 280 mest viste titler (ordnet etter IP-adresser) samlet for bøker og tidsskrift:

Utgivelsesår *	0	* definert som 2007. Ingen verk i tjenesten
1 - 2 år	8 %	
3 - 6 år	12 %	
7 - 11 år	17 %	
> 11 år	62 %	

Bruk etter tema og sjanger - basert på 100 mest viste titler (ordnet etter IP-adresser) samlet for bøker og tidsskrift:

Skjønnlitteratur	9 %	
Faglitteratur	91 %	Herav:
(Kultur)historie, biografier	70 %	
Samfunnsfag, strategiske studier	7 %	
Natur, friluft, reiser	18 %	
Språk, litteratur	9 %	

Antall verk : antall sider vist - basert på 100 mest viste titler (ordnet etter IP-adresser) - samlet for bøker og tidsskrift:

2 verk	5 %
20 verk	25 %
100 verk	66 %

Tjenesten ble lenket til en rekke søkemotorer i januar 2008. Etter dette viste antall besøk/antall sider vist en betydelig økning.

4. Vurderinger

Resultatene av brukerundersøkelsen som blir presentert ovenfor, er en ren kvantitativ kartlegging. Ved pilotprosjektets avslutning tas det sikte på å utarbeide en samlet evaluering som i tillegg vil inneholde en kvalitativ analyse.

Her kommenteres noen hovedfunn i analysen:

- At 78% av bøkene og 83% av tidsskriftartiklene er besøkt, dokumenterer at tjenesten har vært interessant for brukerne. Prosjektgruppen har ikke lyktes med å finne undersøkelser andre steder som fullt ut kan sammenlignes med den aktuelle, for således å kunne vise til en relevant referanse. Ett eksempel som berører brukerfrekvens: "Nearly three-quarters of sessions saw content viewed ..."⁴ Den siterte granskningen analyserer et prosjekt ved University College London, og handler om studenters bruk av pensumlitteratur.
- At 62% av titlene i tjenesten som er besøkt, er utgitt første gang for 11 år siden eller tidligere, kan signalisere at internett er en effektiv distribusjonskanal for litteratur, særlig for slike titler som er vanskelig å få tilgang til på annet vis. Man ser også at det er god spredning i fagområdene det vises interesse for. Det faktum at slike titler stilles til disposisjon, øker også etterspørselen. Og det er delmarkedene eller nisjene som krediteres. Dette fenomenet omtales gjerne som "den lange halen" ("The Long Tail"), et begrep som først ble lansert av den amerikanske skribenten Chris Anderson. Poenget er at flertallet av produkter (f. eks. bøker) som tilbys, har tilnærmet samme verdi som de mest etterspurte titlene, rett og slett fordi de er svært mange og selges over lang tid. Anderson sier: "This is not just a quantitative change, but a qualitative one, too. Bringing niches within reach reveals latent demand for non-commercial content. Then, demand shifts towards the niches [...], creating a positive feedback loop that will transform entire industries – and the culture – for decades to come."⁵ Tallene i herværende analyse synes å vise en klar "hale-effekt".
- Antall brukere og antall sidevisninger økte betydelig da verkene ble søkbare i de store søkemotorene i januar 2008. Dette samsvarer med undersøkelser i andre land. Ett eksempel: "Those people accessing via search engine were most likely to record more views in a session and were more likely to view textpages."⁶
- Tjenestens nettsted inneholder lenker til mer enn 200 nettsted/nettdokumenter med innhold som utgjør kjernestoff om nordområdetemaet. Ettersom disse verkene er født digitalt, havner de ikke i indeksen over digitaliserte nordområdeobjekter, og vil følgelig ikke påvirke tallene i undersøkelsen.
- Tilfanget av bøker og tidsskrifter i tjenesten har ikke vært konstant, men voksende. Den siste suppleringslista (rundt 100 titler) begynte å komme inn i tjenesten i april 2008. De utgjør dermed en ikke ubetydelig del av materialet, samtidig som de er så ferske i tjenesten at besøkstallene for disse verkene er ubetydelige. Dette vil over tid påvirke gjennomsnittstallene for besøk og bruk.
- Undersøkelsene som det vises til ovenfor (note 3 og 5), analyserer to digitaliseringsprosjekter ved akademiske institusjoner i England og USA. Styringsgruppen har studert analysene, men har kommet til at de bare i noen grad er egnet som referanseramme; forskjellen i målgrupper, ulik lengde på prosjektperioder og avvikende analysekriterier tilsier det. På den andre siden kan man likevel konstatere at tendensen og mønstrene i den brukeradferden som dokumenteres, sannsynliggjør et stykke på veg resultatene i herværende analyse.

⁴ David Nicholas et al.: "SuperBook", i programheftet til konferansen "Online Information 2007", London 4. – 6. desember 2007, s. 50.

⁵ Se Chris Anderson: *The Long Tail. How Endless Choice is Creating Unlimited Demand*, Random House 2006, s. 26. Sitatet fremst i denne rapporten er hentet fra Andersons artikkel "Google and the Long Tail of Time", http://thelongtail.com/the-long-tail/2006/04/google_and_the_html, s. 1.

⁶ Nichols, op.cit., s. 54. Se også David Nichols et al.: "What does usage data tell us about users?", i programheftet til konferansen "Online Information 2007", London 4.- 6. desember 2007, s. 84

5. Veggen videre

Et mål med pilotprosjektet har vært å skaffe erfaringstilfang som kan gi impulser til videre samarbeid om digital formidling av rettighetsbelagt materiale.

Undersøkelsen som her presenteres, viser god brukertrafikk i tjenesten og materialet som etterspørres, har betydelig spredning i foretrukne fagområder og en interessant alderssammensetning i forhold til utgivelsesår. Siden en del av innholdet først nylig ble inkludert i tilbudet, har styringsgruppen besluttet å utvide prosjektperioden med 15 måneder. Slik ønsker man også å kvalitetssikre resultatene ytterligere.

Parallelt med ønsket om å foredle pilotprosjektet har styringsgruppen startet arbeidet med å vurdere en utvidelse av samarbeidet. En slik utvidelse vil bli konkretisert når pilotprosjektet evalueres i månedsskiftet oktober/november 2008.