

Bibliotekarstudentens nettleksikon om litteratur og medier

Av Helge Ridderstrøm (førsteamanuensis ved OsloMet – storbyuniversitetet)

Sist oppdatert 06.05.24

Om leksikonet: https://www.litteraturogmedieleksikon.no/gallery/om_leksikonet.pdf

Søkemotor

Også kalt søkemaskin. Et dataprogram som søker gjennom f.eks. Verdensveven (www) og finner flest mulig relevante dokumenter for en bruker. Det kan søkes etter verbal tekst, men også etter bilder, lydfiler m.m.

En søkemotor er avhengig av en type digital robot (blant annet kalt en “crawler”, “wanderer”, “worm” og “spider”) som hver dag “sanker” millioner av nettsider og registrerer sentrale termer på disse nettsidene (innholdsord som “filosofi” og “lymfekreft”, ikke funksjonsord som “hvis” og “her”). Termene blir samlet i en indeks (dvs. en liste eller et register). Søkemotoren kan deretter søke igjennom denne indeksen og finne de termene som en person har skrevet inn i søkemotorens søkefelt. Nettsider med de rette termene blir så lynraskt plukket ut og dukker opp på brukerens skjerm. Disse nettsidene rangeres etter ulike kriterier og kan f.eks. vises med deler av teksten som inneholder det ordet eller de ordene som brukeren har søkt på.

Dirk Lewandowski påpeker i boka *Understanding Search Engines* (2023) at “search engines have an enormous social significance. This can be explained, on the one hand, by their mass use and, on the other hand, by the ranking and presentation of search results. Search engines (like other services on the Internet) are used en masse. Their importance lies in the fact that we use them to search for information actively. Every time we enter a query, we reveal our interests. With every search engine result page (SERP) that a search engine returns to us, there is a (technically mediated) interpretation of both the query and the potentially relevant results. By performing these interpretations in a particular way, a search engine conveys a specific impression of the world of information found on the Web. For every query, there is a result page that displays the results in a specific order. Although, in theory, we can select from all these results, we rely heavily on the order given by the search engine. De facto, we do not select from the possibly millions of results found but only from the few displayed first.” (Lewandowski i <https://link.springer.com/book/10.1007/978-3-031-22789-9>; lesedato 26.04.24)

“Det finnes to typer søkeverktøy på Internett. Søkemotorer og emne kataloger. Søkemotorer bruker roboter som automatisk bygger opp en søkbar indeks av internettsider. De største søkemotorene prøver så godt de kan å lage en mest mulig

komplett oversikt over alle internettsider som finnes i verden. Ingen av dem har hittil klart det, men nærmest til å nå målet er [i år 2000] norske Alltheweb.com som hevder at de nå har lagret informasjon om 300 millioner internettsider i sin indeks. Det er i underkant av 40 prosent av alle internettsider i verden. Emnekataloger er en annen måte å finne fram til internettsider på. Her sorteres informasjonen “for hånd”, og en egen redaksjon lager en hierarkisk oversikt over sider de mener er verdt å katalogisere.” (Larsen 2000 s. 24)

“En søkemotor er en programvare som søker frem informasjon på internett for oss. Det er mange utfordringer forbundet med å finne frem informasjon på nettet. Informasjonen vi forsøker å finne er spredt rundt på servere over hele verden, og i utgangspunktet ville det tatt uendelig lang tid å skanne gjennom alt for å finne det vi ønsker. [...] For å håndtere utfordringer med å finne informasjon på nett, har det blitt utviklet mer og mer effektive søkemotorer, som Google, Yahoo, Bing og mange flere. [...] En søkerobot katalogiserer og indekserer informasjonen i store databaser. Så istedenfor å lete rundt på servere i hele verden, finner man frem i disse indekserte databasene ved hjelp av søkeordene man taster inn. På samme måte som man tidligere søkte i bibliotekdatabaser på det lokale biblioteket etter bøker. Når man har funnet det man leter etter i den indekserte databasen, så har man en direkte adresse til den aktuelle serveren hvor man går ut og henter den relevante informasjonen. På grunn av stadig økende lagringskapasitet og lavere kostnad for maskinvare, lagrer man nå også utsnitt av tekstene i disse databasene. For at man skal få tak i mest mulig relevant informasjon, vil søkemotorene også i forkant rangere de forskjellige dokumentene.” (<https://jtmconsult.no/hva-er-en-sokemotor/>; lesedato 18.01.20)

“Søkemaskinene har blitt våre trofaste veivisere i den etter hvert ufattelige store mengde informasjon som finnes på Internett. [...] Kunsten å mestre søkemaskinene er i dag avgjørende for ikke å gå seg vill på det store nettverket.” (Larsen 2000 s. 11)

Funksjonaliteten i et søkesystem er “[f]eatures built into a search interface that determine the ease with which users may formulate queries and obtain results. Well-designed search software enables the user to:

- Identify the database(s) most appropriate to the subject of the search
- Search more than one database simultaneously
- Remove duplicate records from results when searching multiple databases
- View a current list of the publications indexed in each database
- Choose a search mode based on proficiency (basic or advanced)
- Access well-organized contextual help at each stage of the search process
- View an online thesaurus or use vocabulary mapping to identify appropriate search terms
- Limit search results by relevance, material type, publication date, language, institutional holdings, latest update, full-text, etc.

Sort results by author, title, publication date, etc.
View search terms highlighted in results
Select/deselect records as output
View and save search history
Save results independent of Web browser
Print, e-mail, and download results in various formats” (Joan M. Reitz i http://lu.com/odlis/odlis_c.cfm; lesedato 30.08.05)

“Ønsker du å komme høyt opp i søkeresultatene til Google, gjelder det å lære seg å tenke som en søkemotor. Kun da kan du forstå og kunne tilpasse nettsidene dine i forhold til deres tenkemåte. Problemet i denne sammenheng er at ingen, bortsett fra Google selv, vet hvordan Google virker. Google og de andre søkemotorene gir ut mye informasjon om hvordan deres søkemotor virker, men de holder samtidig enkelte ting hemmelig av konkurransehensyn og for å unngå at folk prøver å skaffe seg en ufortjent plassering. [...] En søkemotor er et nettsted som inneholder en database med informasjon om alle de nettsidene søkemotoren vet om. Ved å angi et søkekriterie kan brukerne her søke i hele denne databasen etter nettsider som oppfyller det oppgitte søkekriteriet. Etter at søket er fullført vises en liste over alle de nettsidene søkemotoren fant som oppfylte det oppgitte søkekriteriet.” (<https://estudie.no/sokemotor/>; lesedato 07.02.20)

Rangeringen i Google bestemmes ikke minst av hvor mange andre nettsider i Googles database som lenker til en bestemt nettside. En nettside som det lenkes til fra sider som er høyt rangert, vil dukke opp langt oppe i søkeresultatet.

“Det vakte bred indignasjon blant researchere og søkemaskin-eksperter da AltaVista i 1999 publiserte planer om å “selge ut” trefflistene sine til kommersielle interesser. Det vil si at et firma f.eks. kan kjøpe et eller flere søkeord i den populære søkemaskinen, for på den måten å betale seg til toppen av trefflistene. I fremtiden kan et søk på ordet “fotballsko” derfor føre til at AltaVista favoriserer produsenten Adidas, og gir denne plass på de ti øverste treffene på lista. En slik utvikling vil selvsagt være uheldig, sett ut fra kravet til at de mest relevante treffene skal listes først.” (Jan Gunnar Furuly i Larsen 2000 s. 13)

“Search engines have been around for many years. For example, MEDLINE, the online medical literature search system, started in the 1970s. The term “search engine” was originally used to refer to specialized hardware for text search. From the mid-1980s onward, however, it gradually came to be used in preference to “information retrieval system” as the name for the software system that compares queries to documents and produces ranked resultlists of documents. There is much more to a search engine than the ranking algorithm [...] Search engines come in a number of configurations that reflect the applications they are designed for. Web search engines, such as Google and Yahoo!, must be able to capture, or *crawl*, many terabytes of data, and then provide subsecond response times to millions of queries submitted everyday from around the world. Enterprise search engines – for

example, Autonomy – must be able to process the large variety of information sources in a company and use company-specific knowledge as part of search and related tasks, such as *data mining*. Data mining refers to the automatic discovery of interesting structure in data and includes techniques such as *clustering*. Desktop search engines, such as the Microsoft Vista search feature, must be able to rapidly incorporate new documents, web pages, and email as the person creates or looks at them, as well as provide an intuitive interface for searching this very heterogeneous mix of information. There is overlap between these categories with systems such as Google, for example, which is available in configurations for enterprise and desktop search.” (Croft, Metzler og Strohman 2015)

“The “big issues” in the design of search engines include the ones identified for information retrieval: effective ranking algorithms, evaluation, and user interaction. There are, however, a number of additional critical features of search engines that result from their deployment in large-scale, operational environments. Foremost among these features is the *performance* of the search engine in terms of measures such as *response time*, *query throughput*, and *indexing speed*. Response time is the delay between submitting a query and receiving the resultlist, throughput measures the number of queries that can be processed in a given time, and indexing speed is the rate at which text documents can be transformed into indexes for searching. An *index* is a data structure that improves the speed of search.” (Croft, Metzler og Strohman 2015)

“Another important performance measure is how fast new data can be incorporated into the indexes. Search applications typically deal with dynamic, constantly changing information. *Coverage* measures how much of the existing information in, say, a corporate information environment has been indexed and stored in the search engine, and *recency* or *freshness* measures the “age” of the stored information. Search engines can be used with small collections, such as a few hundred emails and documents on a desktop, or extremely large collections, such as the entire Web. There may be only a few users of a given application, or many thousands. *Scalability* is clearly an important issue for search engine design. Designs that work for a given application should continue to work as the amount of data and the number of users grow. [...] many different aspects of the search engine, such as the ranking algorithm, the interface, or the indexing strategy, must be able to be tuned and adapted to the requirements of the application.” (Croft, Metzler og Strohman 2015)

“[H]vis et nettsted som Google vet om i dag lager en link til et nytt nettsted eller en ny nettside, så vil Google finne dette nettstedet og denne nettstedet neste gang de scanner nettstedet som har linken [...] Dataprogrammet søkemotorene bruker for å finne og følge lenkene på sidene de vet om kalles for “søkerobot“, “spider” eller “edderkopp“. Tre ord for den samme ting. Foruten å finne og følge alle lenker de finner på en side, benyttes disse søke-robotene til å finne informasjonen som finnes på siden og til å indeksere den i søkemotorenes databaser. [...] Når Google skal

rangere nettsidene i sine søkeresultater (SERP [dvs. search engine results pages]) legger de vekt på følgende hovedkriterier:

Relevans: Relevans vil si hvor relevant Google finner nettsiden i forhold til søkerens geografiske plassering og søkeuttrykk. Dette er den viktigste enkeltfaktoren når Google skal rangere nettsidene [...]

Popularitet: Jo flere nettsider som lager en link til en bestemt nettside, jo mer populær anser Google denne nettsiden og jo høyere vil de rangere nettsiden. Dette fordi de anser mange inngående linker til en bestemt nettside som et kvalitetsstempel. [...] For å få fullt utbytte av inngående linker til ditt nettsted må linkene komme fra nettsteder som selv er høyt rangert i Google på popularitet og de må komme fra nettsteder med høy relevans. Dvs. fra tematisk like nettsider og ikke fra nettsider som ikke har noe med det nettstedet som det linkes til.

Ferskhets: En nettside er en ferskvare. Jo nyere nettsiden er, jo mer relevant anses den også.

Brukervennlighet: Jo mer brukervennlig Google finner en nettside, jo bedre rangerer siden i SERP.

Tillit: Alt som skjer på nettet er basert på tillit. Sprer et nettsted malware, spam eller annet uønsket innhold mister nettstedet automatisk all sin opparbeidede tillit og faller dermed helt ut av Google sine søkeresultater. Å bygge tillit tar imidlertid lang tid og er derfor en vedvarende prosess.” (<https://estudie.no/sokemotor/>; lesedato 07.02.20)

“Er vi i tvil, googler vi. Skal vi ha hjelp til å skrive en CV, googler vi. Ønsker vi å vite mer om hva frihet betyr, googler vi. På bare 15 år har Google blitt en naturlig inngangsport til informasjonshavet som lagres på nettet, hvor det “å google” har blitt et verb på lik linje med det å gå eller sykle. Ifølge Googles oversikt over den norske søketoppen lurte flest nordmenn i 2012 på hva kultur, substantiv eller Instagram er, og flest søkte råd om hvordan få slettet sin Facebook-profil, eller bli gravid. Svaret Google gir på de ulike spørsmålene i de ulike tilfellene, er avhengig av en rekke ulike faktorer. I en rapport fra 1999 beskriver Google-grunnleggerne Larry Page og Sergey Brin for første gang logikken i *PageRank*-algoritmen som selskapet bruker for å organisere og sortere nettsider i sine søkeresultater. Ifølge artikkelforfatterne er relevans, forstått som en nettsides viktighet for en bestemt type søk, et høyst subjektivt anliggende, og avhengig av søkerens personlige interesser, kunnskaper og meninger. [...] Den grunnleggende tanken bak *PageRank* er å rangere et dokument på nettet basert på kvantitet og kvalitet på lenkene som viser til denne nettsiden. Med andre ord, jo flere troverdige kilder som lenker til en bestemt side på nettet, jo viktigere blir denne nettsiden tatt for å være, og dermed plassert høyere opp på resultatlistene i Google. I dag er det over 200 forskjellige faktorer som spiller inn på hvordan algoritmen avgjør Googles svar på hva kultur

er. I tillegg til nettets lenkestruktur bruker Google en mengde tidsspesifikke, geografiske, lingvistiske og søkerelaterte kriterier i sin sortering.” (Taina Bucher i <http://tainabucher.com/wp-content/uploads/2009/08/Bucher-algo-frihet.pdf>; lesedato 06.02.20)

“Man kan [...] tilpasse sin informasjon slik at den blir rangert best mulig. Men søkemotorene har i de senere år utvidet disse tjenestene slik at man nå også kan betale for å få en bedre rangering. Jo mer man investerer, jo bedre rangering kan man få hos den søkemotoren man kjøper denne tjenesten av. Men når man velger å ikke betale lengre, vil man falle tilbake til en vanlig SEO-rangering [dvs. søkemotoroptimalisering, for å rangeres høyere i søkemotorer], som kan ha blitt forbedret av at du har betalt i en periode.” (<https://jtmconsult.no/hva-er-en-sokemotor/>; lesedato 07.02.20)

Søkemotoren Google var blant de første som tillot “og” som logisk operasjon i søkefeltet (“lymfekreft OG pårørende”).

“False drops generally occur when meaning is contingent on the order of search terms (library + school retrieves “library school” and “school library”) or when a term used in a search statement has more than one meaning. For example, a search on the keyword “aids” will retrieve records for items about HIV infection and also items about hearing aids, teaching aids, band-aids, etc. To avoid this problem, a qualifier such as “disease” must be added to the search statement to make retrieval more precise. Synonymous with false combination.” (Joan M. Reitz i http://lu.com/odlis/odlis_c.cfm; lesedato 30.08.05)

I juni måned i 2008 foretok borgere i USA 11,5 milliarder søk på søkemotorer, over halvparten av dem (61,5 %) på Google. I Frankrike var det i mai 2008 ca. 3 milliarder nettsøk med søkemotorer, foretatt av 26 millioner franske internettbrukere, dvs. ca 3,6 søk per dag, og 82 % av dem på Google (<https://journals.openedition.org/questionsdecommunication/592>; lesedato 19.06.20). De fleste brukerne var unge. Personer som er født etter 1993, har blitt kalt “Generation Google”.

“The fact that search engines are a mass phenomenon can also be seen in the number of daily queries. Market research companies estimate the number of queries sent to Google alone at around 3.3 trillion in 2016 (Internet Live Stats & Statistic Brain Research Institute, 2017) – that’s more than a million queries per second! An additional level of consideration arises when we look at how users access information on the World Wide Web. While there are theoretically many access points to information on the Web, search engines are the most prevalent. On the one hand, Web pages can, of course, be accessed directly by typing the address (Uniform Resource Locator; URL) into the browser bar. Then there are other services, such as social media services, which also lead us to websites. But none of these services has achieved a level of importance comparable to that of search

engines for accessing information on the Web, nor is this situation likely to change in the foreseeable future. Last but not least, search engines are also significant because of the online advertising market. The sale of ads in search engines (ads in response to a query) accounts for 40% of the market (Zenith, 2021); in Germany alone, search engine advertising generated sales of 4.1 billion euros in 2019 (Statista, 2021). This form of advertising is particularly attractive because, with each search query, users reveal what they want to find and thus also whether and what they might want to buy. This makes it easy for advertisers to decide when they want to offer their product to a user.” (Dirk Lewandowski i <https://link.springer.com/book/10.1007/978-3-031-22789-9>; lesedato 26.04.24)

“I 2005 innførte Google såkalt *personalisert* søk, søkeresultater som tar høyde for individuell søkehistorikk og signaler relatert til søkerens personlige profil. I praksis betyr det at ingen to søkeresultater er helt like. Det svaret jeg får, er ikke det samme som du får når vi bruker samme søkeord. [...] I tillegg til kode, matematisk-logiske ligninger, instruksjoner og programmerere innebærer det å forstå algoritmer at vi også må undersøke selskapene som utformer algoritmene, deres utviklingsmetoder og ideologi, samt brukere og deres nettverk, for å nevne noen sentrale variabler.” (Taina Bucher i <http://tainabucher.com/wp-content/uploads/2009/08/Bucher-algo-frihet.pdf>; lesedato 06.02.20)

“Den beste og sikreste metoden for å få vekk et treff fra søkemotorer, er ved å få innholdet slettet fra kildesiden, altså nettsiden søkemotoren henter innholdet fra. [...] I 2014 kom EU-domstolen med “retten til å bli glemt”, som pålegger søkemotorene å tilby folk en mulighet til å få fjernet treff om seg selv, selv om kilden til treffet fortsatt finnes. Både Google og Bing har laget egne løsninger for dette [...] Dersom et søketreff har en død link, altså en link som ikke lenger leder til eksisterende innhold, kan man sende en forespørsel til søkemotoren om å få den fjernet eller oppdatert.” (https://min.utdanning.no/utdanningsvalg_artikkel_hvordan_slette_et_treff_fra_en_sokemotor; lesedato 06.03.20)

“Hva søker egentlig folk etter? Det hjelper ikke å være nummer én på “familebil” hvis folk flest søker etter “stor bil”. Søker folk flest etter “slalåmsko” eller “slalåmsko”? Et godt tips er å bruke Google Suggest og se hva Google foreslår. Da ser du også hvilke støtteord du bør ha med. Støtteordene bidrar til at du får bedre rangering. [...] Metadata er informasjon som følger nettsidens adresse, og er helt essensielt for god SEO. Det er denne informasjonen som vises i trefflister som Google og i sosiale medier. En god metabeskrivelse gjør det lettere for brukeren å forstå innholdet, og får flere til å kikke på og dele det. [...] Innholdet på siden skal struktureres etter viktighetsnivå, der tittelen er viktigst. Tittelen må være lik nettadressen, og skal inneholde ordet du ønsker å rangere på. [...] universell utforming for svaksynte [...] Googles bildesøk blir brukt av mange som heller vil scrolle seg gjennom bilder til de finner det som matcher det de leter etter.” (Hilde Erlingsen i <https://www.sempro.no/nyheter/sokemotoroptimalisering-seo>; lesedato 10.09.20)

“Plasseringen som en side får ved opplisting i en søkemotor påvirkes av mange forhold. I noen søkemotorer er det også mulig å betale for en høy plassering. Google henter en stor del av sine inntekter fra denne formen for annonsering, men de har et visuelt skille mellom disse lenkene og de treffene som kommer opp ved ordinært søk. Hvordan rangeres så de treffene hvor det ikke er snakk om å betale seg til en plassering? Rekkefølgen påvirkes blant annet hvor mange ganger ordet det søkes etter forekommer på den aktuelle siden, samt av hvilken metainformasjon som er lagt inn. Metainformasjon er informasjon om sidens innhold, men som ikke vises direkte i nettleseren. Søkemotoren kan også ta hensyn til om de aktuelle ordene er brukt i pekere, og om det forekommer mye brukte kombinasjoner av ord der søkeordet er med. Ikke minst teller sidens popularitet, noe som blant annet kan måles ved å telle hvor mange andre nettsteder som peker til den aktuelle siden. [...] Der Google opererer med Pagerank har Ask.com det de kaller ExpertRank, en metode som skal identifisere de mest autoritative sidene på Internett. ExpertRank skal dermed fastslå popularitet blant de som anses for å være eksperter på et emne. Hva som faktisk fungerer best er imidlertid opp til hver enkelt av oss å avgjøre.” (<http://www.digitalferdighet.no/produsere/soekemotorer>; lesedato 06.03.20)

“Har du noen gang irritert deg over å ikke finne ut hvilken sang du går rundt og nynner på? Nå er løsningen her. Har du en mikrofon, kan du bli kvitt det svært irriterende problemet med å finne ut hvilken sang du hørte på radio, i butikken eller på et utested. Nå holder det bare at du husker nok til å kunne nynne låten, en mikrofon og internett-tilgang for å finne låten igjen. Mange har lenge snakket om det, teknologien har vært der lenge, men det er ikke før nå at du kan bruke din egen stemme som søkeparameter. Det betyr at du kan snakke eller synge inn i en mikrofon, så kan du søke på det du nettopp sa eller sang. Det er søkemotoren Midomi du må finne frem til på nettet hvis du ønsker å søke på din egen nynning. Konseptet til søkemotoren bygger, i tillegg til musikkgjennkjennning, på et sosialt nettverk, så det er brukere over hele verden som legger på tonen til bestemte låter, og som i neste omgang danner et grunnlag for søk, skriver SearchEngineWatch. [...] Teknologien bak søkemotoren er det proprietære Multimodal Adaptive Recognition System (MARS), som er utviklet av Melodis Corporation. Ifølge Melodis, analyserer MARS hvert element uavhengig og inkorporerer de sterkeste komponentene. For eksempel tar MARS-søket tekst med i beregningen hvis brukeren synger teksten, og den ignorerer taleinnhold hvis brukeren nynner eller plystrer låten. [...] Brukere som har opprettet en profil på nettstedet kan synge sine favorittsanger, dele dem med andre og legge dem til i søkemotorens musikk-database. Midomi.com tilbyr også brukerne et system for å gi karakter til andres fremførelse, se på bildene deres og sende dem beskjeder.” (<http://www.aftenposten.no/digital/nyheter/Finn-lata-du-nynner-pa-6445628.html>; lesedato 10.12.13)

Såkalte filterbobler bekrefter de meningene som søkeren allerede har. Mye informasjon blir filtrert bort fordi algoritmene er innstilt på å gjøre den enkelte bruker fornøyd, og det avhenger av kva brukeren har søkt på tidligere, brukerens

kjønn, alder m.m. “Når du søker på Google, sørger en algoritme for å gi deg informasjon som stemmer med de synspunktene som inngår i brukerprofilen din. Slik isoleres vi i vår egen kulturelle og ideologiske boble. [...] Tilpassingen skjer ut fra innsamlet informasjon om brukernes preferanser. Det gjelder både kommersielle tjenester, nyhetstjenester og søketjenester. Et godt eksempel er nettbokhandelen Amazon. De vet hva kunden tidligere har kjøpt, og hva andre, som har de samme interessene, har kjøpt. Denne personifiseringen er en bevisst strategi hos de største internett-selskapene som Facebook, Google, Microsoft og Yahoo. Men en ting er å samle informasjonen for å selge en vare. Det er noe helt annet når denne informasjonen brukes for å styre vår tilgang til informasjon. [...] Vi beveger oss mot en verden der Internett viser oss det de tror vi ønsker å se, men ikke nødvendigvis det vi trenger å se, ifølge Eli Pariser. Han mener at ved å kombinere alle disse filtrene, alle disse algoritmene, så får vi det han kaller for en filterboble, vårt eget begrensede informasjonsunivers. Problemet er at vi ikke selv bestemmer hva vi vil ha av informasjon, og at vi ikke vet hva som blir filtrert bort.” (<https://ndla.no/nb/subjects/subject:14/topic:1:185993/resource:1:107782>; lesedato 10.09.20) Begrepet ble skapt av den amerikanske aktivisten og forfatteren Eli Pariser, og gjort kjent gjennom boka *The Filter Bubble: What The Internet Is Hiding From You* (2011).

“Du legger igjen spor etter deg hver gang du bruker internett. Google, for eksempel, “lærer seg hvem du er” gjennom å ta vare på informasjon om deg som: hva du pleier å søke på, hva du pleier å gjøre på nett og den henter personlig informasjon du legger fra deg om hvem du er på forskjellige steder. Det er det de kaller “big data”. Du har gjort noen valg og da er det dataspor som blir lagret til en annen gang, eller selges til andre. Det er avslørt en sjokkerende mengde personlig informasjon om oss, slik som våre interesser, familieforhold, legning, venner, bekjente, politisk orientering, medisinske tilstander, kortinformasjon og mer. Denne informasjonen er nåtidens “olje” for markedsførere, ondsinnede hackere og kriminelle. Alle sammen ville sette stor pris på å få fatt i dine private søkedata. Noen aktører vet mer om deg enn andre. I dag lever vi i en verden der digitalt medieinnhold mer eller mindre er personalisert, det vil si at informasjon tilpasses deg og meg basert på våre dataspor. Det er for eksempel bevist at ved det amerikanske presidentvalget i 2016 ble det gitt politisk reklame direkte til brukere på nett på bakgrunn av slike data. [...] Fordi vi nå bruker mobiltelefoner og ikke bare datamaskiner, pønsker eksempelvis Google fort ut om du reiser samme sted hver dag, og antar derfor at du jobber på dette stedet. Dette kommer i tillegg til den informasjonen du eventuelt aktivt har forsynt Google med ved å fylle ut din eventuelle profil. Om du ikke har fylt ut dette, gjør Google en kvalifisert gjetning basert på aktiviteten din. [...] Vi får se det algoritmene, altså de matematiske beregningene, tror vi vil se på bakgrunn av våre tidligere valg og søk, og ikke nødvendigvis det vi trenger.” (Marianne Hagelia i <http://www.digitalstudie.teknikk.no/2018/01/23/vi-lever-i-en-filterboble/>; lesedato 19.02.20)

“Det er algoritmer som gjør at Google vet at du søkte på sommerkjole og pyramidene. Det er også algoritmer som sier at “andre som har kjøpt dette produktet – har også kjøpt disse produktene” når du er inne i en nettbutikk. Det er algoritmer som bestemmer om jeg får se feriebildene til min venninne på Facebook eller fisketuren til min fetter, ofte ut i fra hvem jeg har mest med å gjøre gjennom interaksjon som chat, kommentarer og likes. Det er også algoritmene som bestemmer hvor relevant noe er, dvs. hvilken informasjon som til enhver tid fortjener vår oppmerksomhet. Vi kan si at algoritmer er produktive, i den forstand at de er med på å konstruere en bestemt framstilling av verden. En framstilling som er nøye planlagt gjennom klart definerte beregninger. Dette er et økende demokratisk problem. Et demokratisk frihetsbegrep innebærer tilgang på relevant informasjon for å kunne ta rasjonelle valg og felles beslutningsprosesser. Flere og flere roper nå et varsku for både demokrati og våre muligheter for kunnskaps-tilegnelse. De påpeker at ved å kombinere alle disse filtrene, alle disse algoritmene, så får vi det vi kaller for en filterboble, vårt eget begrensede informasjonsunivers. Det er problematisk at vi ikke selv får bestemme hva vi vil ha av informasjon, og at vi ikke vet hva som blir filtrert bort. Algoritmene har blitt de nye portvaktene for informasjon laget av uidentifiserbare programmerere. De erstatter tradisjonelle portvakter som redaktører og journalister, som tross alt er identifiserbare personer. I en verden der informasjon tilpasses og justeres hver enkelt, finnes det noen garanti for at vi har et felles utgangspunkt for en demokratisk diskusjon?” (Marianne Hagelia i <http://www.digitalstudieteknikk.no/2018/01/23/vi-lever-i-en-filterboble/>; lesedato 19.02.20)

“Søketjenestene lagrer informasjon om hva vi gjør, og hva vi liker. [...] Hva du liker av mat, om du er religiøs, har barn, eller bestemte helseproblemer, er ting søkemotorene kan kartlegge. “If the product is free, you are the product”. Det er en frase som går igjen blant nettekspertene som prøver å fortelle at du betaler for “gratisøk” ved å gi bort informasjon om deg selv i stedet for penger. I tillegg til å lagre søkepreferanser, tar selskapene vare på grunnleggende informasjon om IP-adresse, språk og hvor du befinner deg, gjerne så lenge som et års tid. Men det finnes alternativer. [...] Søkemotoren DuckDuckGo fikk en eksplosiv vekst i antall søk [...] rundt fire millioner daglige treff. Det er riktignok mikroskopisk sammenlignet med Google, som hadde et gjennomsnittlig daglig tall på nesten 6 milliarder i 2013. Tjenesten profilerer seg med å tilby universelle søk, som gir det samme resultatet for alle brukere, og som ikke lagrer noe informasjon om søkeaktivitet. Tjenesten skal være helt anonym og kryptert, og lagrer ikke informasjonsskapsler, eller cookies, på datamaskinen din. DuckDuckGo skiller seg dessuten fra Google med å ha en algoritme som vektlegger kvalitet fremfor kvantitet. Informasjon fra sider som er rangert som troverdige blir prioritert, fremfor å hente informasjon fra flest mulige kilder. Søkeresultatene kommer i hovedsak fra brukergenererte sider som Wikipedia, og fra andre søkemotorer som DuckDuckGo har partnerskap med, blant annet Wolfram Alpha, Bing og Yahoo.” (Sigurd Ø. Flæten i <https://www.tu.no/artikler/her-er-sokemotorene-som-gir-deg-andre-treff-og-bevarer-anonymiteten-din/223377>; lesedato 09.12.20)

“Ixquick skryter av å være verdens mest private søkemotor. Den er en såkalt metasøkemotor, som kun henter resultater fra andre motorer. Ixquick bruker et rangeringssystem med stjerner, hvor et enkelt treff får en stjerne hvis en av søkemotorene har truffet som et av sine topp ti resultater. Hvis treffet har fem stjerner, betyr det at fem søkemotorer har det blant sine topp ti treff. Ixquick har en enkel og grei personvernserklæring: “Ixquick samler eller deler ingen personlig informasjon.” Søket krypteres, og at søkepreferansene lagres i kun én cookie, som slettes etter 90 dager. [...] Startpage ser helt lik ut som Ixquick, og leveres av det samme nederlandske selskapet. Forskjellen er at Startpage promoteres som “Enhanced by Google”, som i praksis betyr at den er en anonymisert portal til Google-søk. Google kan dermed bare se at det kommer søk fra Startpages servere, uten å kunne knytte dem til deg. [...] Blekko lagrer riktignok informasjon om deg når du søker, som IP-adresse, hvor du befinner deg, og hvilket språk du bruker. Men denne informasjonen lagres kun i 48 timer, som er betraktelig kortere enn de store, kommersielle tjenestene, som i noen tilfeller lagre dette i opptil 18 måneder før det blir anonymisert. I tillegg kan Blekko tilby en unik søkefunksjon som heter “slashtags”. Slashtags er emneord du kan legge til søket ditt, som gjør at søket begrenses til et visst antall sider. Hvis du for eksempel søker “diabetes /health”, har Blekko selv valgt ut et antall nettsider knyttet til slashtaggen /health, som anses for å være troverdige. Blekko har laget hundrevis av slike slashtags, og det er enkelt å lage egne slashtags for filtrerte søk. [...] Mange vil nok oppleve at de alternative søkemotorene gir en annen type, og kanskje mindre presise treff, enn Google. Da blir avveiningen om man er villig til å gi fra seg personlig informasjon for et mer treffsikkert søk. Mange har svart ja på det spørsmålet allerede.” (Sigurd Ø. Flæten i <https://www.tu.no/artikler/her-er-sokemotorene-som-gir-deg-andre-treff-og-bevarer-anonymiteten-din/223377>; lesedato 09.12.20)

“[S]earch engines, news aggregators, and social networks are increasingly personalizing content through machine-learning models (Agichtein, Brill, and Dumais 2006; Das et al. 2007; Hannak et al. 2013), potentially creating “filter bubbles” (Pariser 2011) in which algorithms inadvertently amplify ideological segregation by automatically recommending content an individual is likely to agree with. Moreover, individuals are more likely to share information that conforms to opinions in their local social neighborhoods (Moscovici and Zavalloni 1969; Myers and Bishop 1970; Spears, Lea, and Lee 1990; Schkade, Sunstein, and Hastie 2007). If realized, such information segregation is a serious concern, as it has long been thought that functioning democracies depend critically on voters who are exposed to and understand a variety of political views (Downs 1957; Baron 1994; Lassen 2005). On the other hand, Benkler (2006) and others have argued that increased choice and social networks lead to greater exposure to diverse ideas, breaking individuals free from insular consumption patterns (Obendorf et al. 2007; Goel, Hofman, and Siner 2012).” (Seth Flaxman, Sharad Goel og Justin M. Rao i <https://5harad.com/papers/bubbles.pdf>; lesedato 10.09.20)

“I sin nyeste bok, *Everybody Lies*, bruker forfatteren Seth Stephens-Davidowitz google-søk som datamateriale for å si noe om amerikanernes skjulte og svært private vaner. Han omtaler Google som det digitale sannhetsserumet og kan skilte med mange interessante funn etter fem års leting. Skal man tro google-søkene, er nivået av rasisme i USA høyere enn hva man har trodd, det er flere homofile enn hva nasjonale statistikken tilsier, og stadig flere kvinner googler hvordan de kan utføre abort på egen hånd. Stephens-Davidowitz har doktorgrad i økonomi fra Harvard og var tidligere forsker ved Google.” (*Morgenbladet* 7.–13. juli 2017 s. 38)

Safiya Umoja Nobles bok *Algorithms of Oppression: How Search Engines Reinforce Racism* (2018) viser at “negative biases against women of color are embedded in search engine results and algorithms [...] Run a Google search for “black girls” – what will you find? “Big Booty” and other sexually explicit terms are likely to come up as top search terms. But, if you type in “white girls,” the results are radically different. The suggested porn sites and un-moderated discussions about “why black women are so sassy” or “why black women are so angry” presents a disturbing portrait of black womanhood in modern society. In *Algorithms of Oppression*, Safiya Umoja Noble challenges the idea that search engines like Google offer an equal playing field for all forms of ideas, identities, and activities. Data discrimination is a real social problem; Noble argues that the combination of private interests in promoting certain sites, along with the monopoly status of a relatively small number of Internet search engines, leads to a biased set of search algorithms that privilege whiteness and discriminate against people of color, specifically women of color. Through an analysis of textual and media searches as well as extensive research on paid online advertising, Noble exposes a culture of racism and sexism in the way discoverability is created online. As search engines and their related companies grow in importance – operating as a source for email, a major vehicle for primary and secondary school learning, and beyond – understanding and reversing these disquieting trends and discriminatory practices is of utmost importance. [...] *Algorithms of Oppression* contributes to our understanding of how racism is created, maintained, and disseminated in the 21st century.” (<https://nyupress.org/9781479837243/algorithms-of-oppression/>; lesedato 14.01.20)

“25. juni 2009, klokken 2:26 lokal tid, begynte ryktet å gå om at popkongen Michael Jackson hadde dødd på UCLA Medical Center i Los Angeles, USA. I løpet av kort tid strømmet millioner av mennesker til Google News for å søke etter informasjon som kunne bekrefte eller avkrefte ryktet. Overraskelsen var stor da de fikk opp denne meldingen i stedet: “Vi beklager, men søket ditt ligner automatiserte forespørsler fra datavirus eller spyware. For å beskytte våre brukere kan vi ikke behandle ditt spørsmål nå.” Dette kom som en overraskelse på dem som forventet å få opp ferske og relevante nyhetsartikler. Så hvorfor kom denne beskjeden i stedet? Jo, fordi Google tolket den enorme flommen av søk på de samme ordene “Michael Jackson død” som et dataangrep. Det fikk søkemotorens

algoritmer til å reagere slik de var programmert til under fiendtlige angrep: de avviste søkene. Det tok 25 minutter før Google fikk ordnet opp.” (Steffen Viken Valvåg og Randi Merete Solhaug https://uit.no/nyheter/artikkel?p_document_id=424614; lesedato 20.10.20)

Historikeren og krimforfatteren Hans Olav Lahlum ble intervjuet i 2016. Intervjueren kommenterte: “- Da jeg googlet deg før vi møttes, kom det opp automatisk på skjermen “Hans Olav Lahlum og IQ”. Det kom også “Hans Olav Lahlum og Asperger”. - Og “diagnose” og “homofil”. Jeg vet det. Jeg har aldri vært til medisinsk utredning, har ingen diagnose. Ikke er jeg homofil heller. Dette kommer opp på Google fordi folk spekulerer om meg og visse egenskaper, sykdommer eller tilstander, og søkemotorer er lagt opp slik at visse kombinasjoner av ord kommer opp hvis mange nok har søkt på den samme ordkombinasjonen.” (*Bok og bibliotek* nr. 4 i 2016 s. 61)

En såkalt metasøkemotor søker via mange søkemotorer.

Litteraturliste (for hele leksikonet): <https://www.litteraturogmedieleksikon.no/gallery/litteraturliste.pdf>

Alle artiklene i leksikonet er tilgjengelig på <https://www.litteraturogmedieleksikon.no>