

Pasientjournal, kunstig intelligens og språkteknologi

Michael 2023; 20: Supplement 31: 87–96.

Etter hvert som helsetjenestene blir digitalisert, endrer pasientjournalens funksjon seg fra hovedsakelig å være en intern dokumentasjon i det enkelte sykehus eller legesenter, til å bli en plattform for kommunikasjon og digitale beslutningssystemer. Journalen er ikke lenger bare helsearbeiderens verktøy til å følge opp pasienten, men datamaskinens verktøy for kommunikasjon og beslutningsstøtte.

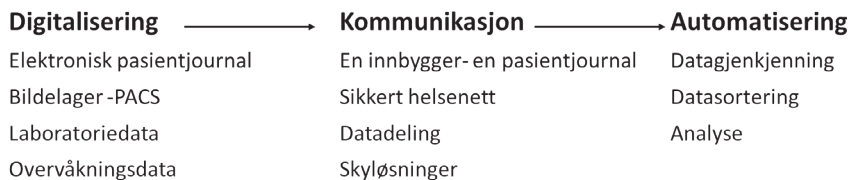
Endringene i bruk av pasientjournalen kan deles i tre faser (figur 1).

1. Digitalisering av all informasjon
2. Bruk av journalen til kommunikasjon mellom helsearbeidere og mellom helsearbeidere og pasienter: *Én innbygger – én pasientjournal*
3. Datamaskinen utfører autonome analyser av tekst og data

Digitaliseringen startet i 1980-årene og er stort sett gjennomført på de fleste samfunnsområder. Men utviklingen har gått skuffende langsomt når det gjelder helsepersonells tilgang til samme pasientinformasjon i alle ledd i helsetjenesten. Det virkelige potensialet for digitalisering av alle typer pasientdata henter vi først ut når vi utnytter datasystemenes analysekapasitet. Der er vi fortsatt på forskningsstadiet.

Å bruke kunstig intelligens til å analysere data i journalen og utføre autonome oppgaver, som beslutningsstøtte, fungere som samtalerobot i kommunikasjon med pasienten, analysere trender og endringer i en pasients tilstand etc., er nylig vist i en rapport fra Teknologirådet (1). I Norge er det gjort arbeid med såkalt naturlig språkanalyse av ustrukturert journaltekst. Men av personvern hensyn har det vært vanskelig å få bruke pasientjournalene til dette (2). I 2021 ble helsepersonelloven endret for å gjøre det enklere å bruke pasientdata til utvikling av kunstig intelligens (3). Dette vil forhåpentligvis stimulere utviklingen av slike løsninger.

Digitalisering i helsetjenesten



Figur 1. Utvikling av digital pasientjournal skjer i tre faser. Første trinn (digitalisering av alle pasientdata inkludert tekst) er snart komplett i Norge. Andre trinn (kommunikasjon) har vært i innføringsfasen siden 2011, da Stortinget i bestemte at journaldata om hver pasient skulle være tilgjengelig for alle enheter i helsetjenesten. Det tredje stadiet (automatisering) vil si at vi ved hjelp kunstig intelligens utvikler programmer som kan analysere data i journalen og utføre autonome oppgaver som beslutningsstøtte, fungere som samtalerobot i kommunikasjon med pasienten, analysere trender og endringer i en pasients tilstand, etc.

Hvorfor tar det så lang tid?

Pasientjournalen består hovedsakelig av ustrukturert tekst, der det er stor variasjon både i bruk av fagtermer og norsk språk. Dette kan fungere for helsepersonell som er vant til dette, men datamaskinene arbeider lettest med strukturerte data og entydige fagtermer og mindre variasjon i språket.

For å gjøre journaldata bedre tilgjengelig for digital behandling, er det derfor utviklet felles nomenklaturer. I dag er SNOMED CT et utbredt nomenklatur- og kodeverk som forvaltes av *International Health Terminology Standards Development Organisation* (IHTSDO). Land kan melde seg inn og ta i bruk systemet. Den norske versjonen av SNOMED CT var klar i 2021.

Direktoratet for e-helse har lagt SNOMED CT til grunn for terminologien i Norsk helsenett og til utvikling av målbildet *En innbygger – én pasientjournal*, som Stortinget vedtok i 2011. Man har prioritert arbeidet med å innføre denne nomenklaturen og å innføre en helseplattform som prioriterer datamaskinenes behov (4).

Jeg vil her presentere noen erfaringer med å utnytte den ustrukturerte journalteksten til både autonom pasientkommunikasjon og analyse.

Thesaurusprosjektet

I 2000 var *Doculive* (Siemens) som elektronisk journalsystem i bruk på flere avdelinger ved Rikshospitalet. *Doculive* ble brukt til å skrive og lagre jour-

nalen, men hvis vi ville dele den med pasienten, måtte vi skrive ut en papirkopi. På grunn av økende etterspørsel om å få innsyn i journalen, utviklet vi et oversetterprogram som skulle gjøre journalen forståelig for pasientene, uten hjelp av en fagperson. Pasienten skulle få en digital versjon av journalen. Når pasienten førte musepekeren over medisinske ord, kom det opp en norsk oversettelse. Hvis de klikket på en diagnose, for eksempel angina pectoris, fikk de opp en artikkel om diagnosen fra *Store medisinske leksikon* (Kunnskapsforlaget). En bokmålsordliste og tre medisinske kilder ble koblet til journalen: *Felleskatalogen*, *Norsk elektronisk legehåndbok* (Norsk Helseinformatikk) og *Medisinsk ordbok* (Kunnskapsforlaget). For å identifisere ord som ikke ble gjenkjent i noen av oppslagsverkene, lagde vi et program med fargekoder. Alle ord som programmet gjenfant i bokmålsordlisten, forble svarte, mens ordene som programmet gjenfant i de tre medisinske oppslagsverkene, ble grønne. Ord som ikke fantes i noen av ordlistene, ble røde. Hvis forklaringen var forskjellig i de to medisinske oppslagsverkene, beholdt vi begge. Alle ord som programmet ikke gjenkjente og som kom opp i rødt, ble manuelt oversatt. Hvis det var vanlige norske ord som var stavet feil, la vi dem inn som synonymer til riktig stavet ord.

Thesaurusprogrammet ble både et verktøy for å oversette journaltekst og et verktøy til å oppdatere medisinsk ordbok.

Det var flere utfordringer knyttet til den ustrukturerte teksten i journalene:

- Journalspråket er dynamisk og inneholder mange ord og akronymer ingen ordbokredaktør kunne forestilt seg, der fremmedordene kan være både latinskliknende og engelsklignende. For eksempel når en hjertekirurg syr en anastomose mellom venstre arteria intrathoracica interna (på engelsk internal mammary artery) til ramus interventricularis anterior av arteria coronaria sinistra (på engelsk left anterior descending artery), vil han skrive IMA til LAD i journalen.
- Selv om legen skriver inn et latinsk ord, kan det være stavet på forskjellige måter.
- Mange organer i kroppen har samme navn. For å forstå hvilket organ det er snakk om, må vi vite noe om sammenhengen. Ventrikkel i magen betyr magesekk, ventrikkel i hjertet er hjertekammer og i hjernen er ventriklene de naturlige hulrommene.

Da programmet kunne oversette så godt som alle ord i de hjertekirurgiske journalene ved Rikshospitalet, testet vi det på fem hjertekirurgiske pasienter. Pasienten ble bedt om å skalere forståelsen av teksten i egen journal med og uten oversetterprogrammet. Det viste seg at de ikke forstod så mye mer,

til tross for at en ren oversettelse av ordene gjorde at de kunne lese teksten. Pasientene var ikke interessert i teksten i seg selv, men hva innholdet betydde for dem.

Dette kan illustreres ved at vi viste pasientene to forklaringer av akronymet for transitorisk iskemisk angrep, TIA.

1. TIA – transitorisk iskemisk angrep, forbigående nedsatt blodstrøm til en del av hjernen og med forbigående tap av kropps- eller mental-funksjoner, tilstanden normaliserer seg i løpet av 24 timer.
2. TIA – transitorisk iskemisk angrep, forbigående anfall av oksygenmangel i deler av hjernen.

Alle pasientene forsto forklaring 1, men forklaring 2 ble ikke forstått. Pasientene ble også forvirret av at negative funn var positivt for dem.

Thesaurusprosjektet viste at det er mulig å lage programmer som gjør journalteksten forståelig for pasientene, men det er en krevende prosedyre som ikke bare handler om å forstå hva ordene betyr, men hva innholdet som helhet betyr for pasientens helse og prognose (5).

Da sykehuset skiftet journalsystem til DIPS i 2014, måtte vi stoppe thesaurusprosjektet, og Kunnskapsforlaget overtok rettighetene til programvaren. I dag ville vi antagelig ha utviklet mer sofistikerte algoritmer for å automatisere oversettelsen av fremmedord, for eksempel en algoritme for å «oversette» teksten til SNOMED-nomenklatur.

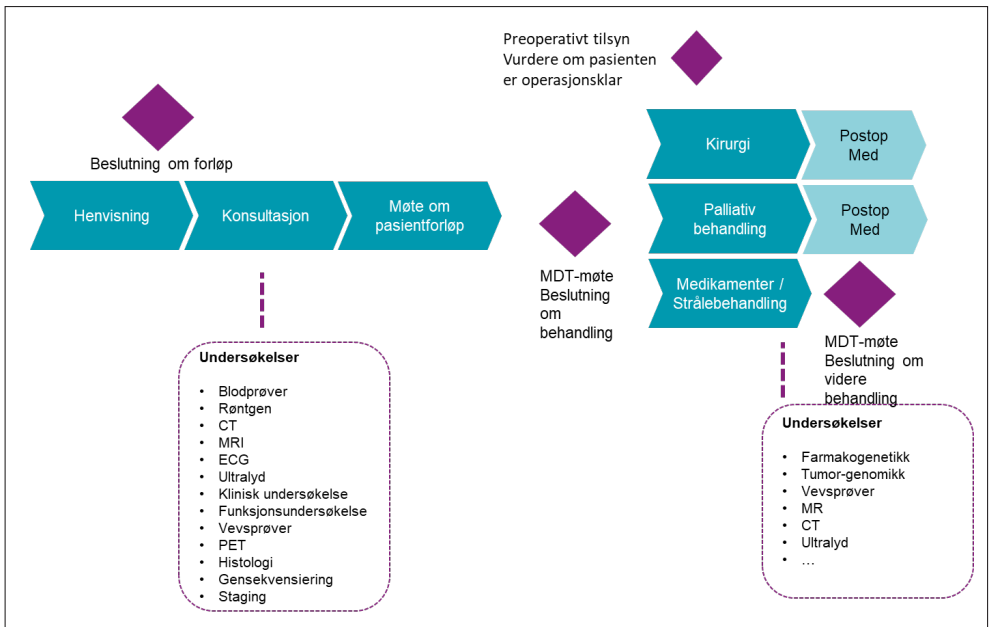
BIGMED-prosjektet


I 2016 fikk vi midler fra Forskningsrådet til prosjektet BIGMED, der målet var å bruke kunstig intelligens til å identifisere og fjerne noen av flaskehalsene for å innføre presisjonsmedisin i klinisk praksis.

Presisjonsmedisin, som også kalles persontilpasset medisin, innebærer at vi skal kunne nyttiggjøre oss all mulig informasjon om en pasient og vedkommendes sykdom, slik at vi kan velge en skreddersydd behandling for akkurat det individet med akkurat den sykdommen.

En av flaskehalsene som hindret en effektiv prosess, var møtene i multi-disiplinære team (MDT), som er kjernen i beslutningsprosessen rundt hver pasient. Informasjonstilfanget blir stadig mer detaljert, og møtene este ut fordi eksperter på alle aktuelle områder burde være til stede. Deltakerne i MDT-møtet skal vurdere og ta beslutning om tre forhold: Eksakt diagnose, de beste behandlingsalternativer og pasientens tilstand slik at pasienten får en behandling som både er effektiv og tolerabel (figur 2).

En av visjonene i BIGMED-prosjektet var at journalen og alle prøvesvarene skulle være tilgjengelige på møtet via en datamaskin med en algoritme



Figur 2. Beslutninger i pasientforløpet i sykehus
 Når pasienten henvises, besluttes forløp. Pasienten blir enten innkalt til poliklinikk eller innlagt. Når alle prøvesvar foreligger gjennomføres et multidisiplinært team-møte (MDT-møte), der man tar beslutning om behandling. Hvis pasienten skal opereres, gjøres det en ny vurdering av om pasienten er klar for inngrepet dagen før. Hvis pasienten trenger ytterligere behandling, gjentas beslutningsprosessen. I alle beslutningene merket med  kan kunstig intelligens som analyserer journalteksten, sikre en riktig beslutning.

som kunne lese og systematisere tekstfiler og prøvesvar, og etter hvert bli med å vekte de ulike informasjonene mot hverandre. Derved kunne MDT-møtene bli mindre personavhengige og mer standardiserte.

Naturlig språkanalyse av ustrukturert tekst

For å få til en slik funksjon, må datamaskinen kunne lese og forstå den ustrukturerte teksten i journalen. Naturlig språkanalyse (natural language processing, NLP) av ustrukturert journaltekst ble derfor en av arbeidspakene i prosjektet. Dette er et område IT-miljøene både ved Norges teknisk-naturvitenskapelige universitet og Universitetet i Oslo har arbeidet med i lang tid. Forskere fra begge miljøene var derfor med i prosjektet.

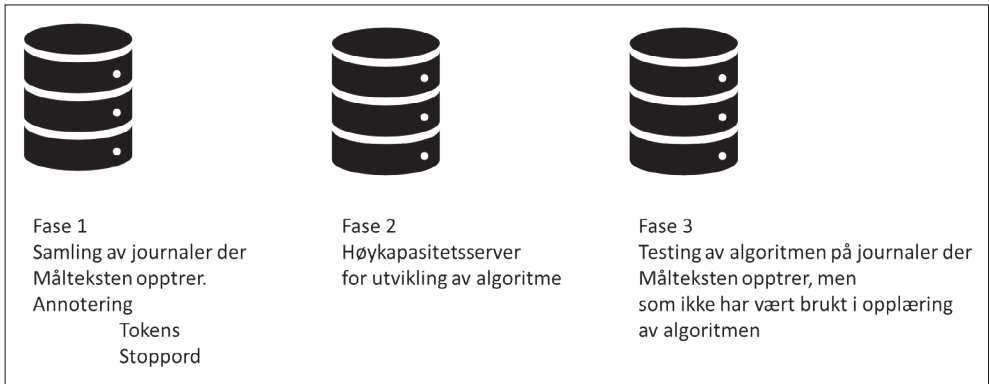
Lilja Øverlid ved Institutt for informatikk, Universitetet i Oslo ledet prosjektgruppen, og Pål Haugar Brekke ved Hjertemedisinsk avdeling, Rikshospitalet var ansvarlig for det kliniske innholdet.

Utvikling av et program eller modell som kan gjenfinne informasjon i naturlig språk består, grovt forenklet, av tre faser. I første fase må vi samle flest mulige journaltekster der slikt innhold (diagnoser, prosedyrer, handlinger og medikamenter) finnes. Så må innholdet annoteres, dvs. at eksperter markerer elementer i teksten, gir dem ulike klasser og definerer relasjoner mellom dem. Disse termene kalles på dataspråket *tokens*, det vil si termer som skal forklares (annoteres) for maskinen, slik at den kjenner dem igjen og derved kan tolke den ustruktureerte teksten. Det er altså en type gjenkjenningsord som kan bestå av enten ett eller flere ord. For å bygge gode modeller, må datagrunnlaget inneholde ikke bare enkeltord som i en ordliste, men språket i en kontekst, og med et bredest mulig utvalg varianter. En grunnleggende hypotese i språkteknologien er at ord som opptrer i samme kontekst, tenderer til å ha samme betydning, og nettopp dette gjør at språkmodeller kan «se forbi» stavfeil, markere synonyme termer osv. De vanlige norske ordene som binder setningene sammen, kalles *stoppord*, og skal ikke ha betydning for algoritmen. På samme måte som pasientene som trengte en kontekst for å forstå den oversatte teksten i thesaurusprosjektet, må maskinen ha en kontekst. I andre fase mates de annoterte dataene inn i en høykapasitets arbeidsstasjon eller beregningsserver som utvikler modellen eller programmet. I siste fase testes modellen på for maskinen ukjent tekst, for å se om den kan markere de interessante elementene og angi en sannsynlighet for hvilken klasse eller betydning de har. Det kan innebære at modellen presenteres for et sett journaler der noen inneholder diagnosen eller tilstanden den skal finne. Så registreres det hvor mange av tekstene diagnosen opptrer i og maskinen finner den (sensitivitet) og i hvor mange den tror diagnosen ikke finnes og der den ikke opptrer (spesifisitet) (figur 3).

Naturlig språkanalyse av familieanamnesen i journalen

Siden arvelige sykdommer og genetiske data var en viktig del av BIGMED-prosjektet, var planen å trene algoritmer på pasienter med mulig arvelig hjertesykdom for å finne sammenheng mellom familieanamnese i journalen og pasientens genotype.

Det tok tid å få tillatelse fra personvernombudet til å bruke journaldata til å trene algoritmen. En kardiolog med lang erfaring med denne pasientgruppen utviklet derfor syntetiske familiehistorier som grunnlag for å trene algoritmen.



Figur 3. Utvikling av naturlig språkanalyse

Basert på de syntetiske familiehistoriene fikk de et korpus på 477 setninger og 6030 tokens, som ble brukt til å utvikle en modell for gjenkjenning av termer og relasjoner. Modellen som var trent på syntetiske data, ble så testet på familieanamneser fra ekte journaler. Det viste seg at modellen kunne gjenfinne termer i de ekte journaltekstene med god presisjon, selv med et begrenset datagrunnlag (6).

Naturlig språkanalyse av røntgenbeskrivelser

I et annet prosjekt utført av samme gruppe i samarbeid med Akershus universitetssykehus, utviklet man en algoritme for å kvalitetssikre CT-undersøkelser av barn. Siden det er en kjent assosiasjon mellom gjentatte CT-undersøkelser av barn og tumorutvikling senere i livet, er det viktig at det utføres færrest mulig slike undersøkelser der det ikke foreligger patologi. Det var derfor av interesse å utvikle kunstig intelligens som kunne klassifisere CT-beskrivelser, ut fra om man hadde gjort patologiske funn.

De utviklet tre forskjellige algoritmer som ble trent på merkede data, der alle beskrivelser som inneholdt beskrivelse av patologiske funn, ble merket som patologiske. De brukte alle undersøkelser av barn gjort i en 12-års periode, totalt 13 506 CT-undersøkelser, og la til 1000 tilfeldige CT-undersøkelser av voksne og 100 fra thoraxrøntgenundersøkelser av voksne. 10 % av undersøkelsene ble tatt ut for å teste algoritmen, og disse ble ikke brukt i utviklingen av algoritmen. Alle de tre algoritmene viste høy sensitivitet og spesifisitet og vil kunne brukes til å kvalitetskontrollere bruk av CT-undersøkelser hos barn i fremtiden (7).

Naturlig språkanalyse for å finne ut om pasienten har allergi

Universitetet i Agder og Sørlandet sykehus utviklet en algoritme som kunne analysere journalene til pasienter som skulle opereres, og fange opp om det står noe i tidligere journaler om allergi. Dette er en viktig problemstilling når vi skal vurdere pasienter for kirurgisk behandling eller for antibiotika og annen medikamentell behandling (se figur 2). Istedenfor å annotere manuelt, brukte de maskinlæringsalgoritmer for å analysere store mengder kliniske historier til å bygge en språkmodell med ord og fraser som data-maskinen kan lære hva betyr.

De fikk ta ut en kopi av all klinisk dokumentasjon på alle innlagte pasienter siden 1992, totalt over 39 millioner dokumenter. Ved å søke etter ord som inneholdt noe om allergier, fikk de annotert ulike kategorier av tokens som inneholdt informasjon knyttet til allergier. Programmet de utviklet, viste seg å kunne identifisere pasienter med allergi i sykehistorien med god presisjon (8).

Kunstig intelligens kan finne nøkkelinformasjon

Disse tre eksemplene viser at det er mulig å bruke kunstig intelligens til å finne nøkkelinformasjon om pasientene i ustrukturert fritekst i norsk pasientdokumentasjon. I årene som kommer, vil det komme mange tilsvarende prosjekter der det utvikles algoritmer som kan finne ulike diagnoser i norske journaler. Det er viktig å samle og koordinere disse prosjektene hvis vi skal få modeller som kan spille en rolle i beslutningene om behandling, slik vi planla i BIGMED-prosjektet. Dette bør være en oppgave for norske helsemyndigheter.

I USA har helsedepartementet prioritert en slik løsning. Food and Drug Administration (FDA) og Center for Disease Control (CDC) fikk i 2016 ansvar for å samle alle erfaringer med naturlig språkanalyse. Målet var å etablere en nettbasert tjeneste der forskere kunne laste ned algoritmer for å strukturere ustrukturerte journaldata og derved kunne sammenligne sykehistorier (9).

De begynte med å få en oversikt over alle språkanalyseprosjektene som pågikk basert på ustrukturert tekst i pasientdokumentasjon i USA. I 2017 hadde de funnet 71 forskjellige språkanalysesystemer som var utviklet på engelsk for å strukturere ustrukturert klinisk tekst. I 35 artikler var naturlig språkanalyse brukt til å analysere journaltekst, i 11 artikler radiologirapporter, i ni artikler patologirapporter, i sju tilfeller biomedisinsk litteratur og i fem kliniske utprøvinger. Bare et fåtall av artiklene var basert helt på maskinlæring uten noen form for manuell læring (10).

I Norge virker det som arbeidet med å tilgjengeliggjøre pasientens journal i alle ledd av helsetjenesten har handlet mer om at helsepersonell skal lære seg et nytt språk, enn å legge forholdene til rette for at helsetjenesten i samarbeid med akademiske miljøer og kommersielle aktører utvikler pasientjournalen til et verktøy for datamaskinene.

Oppsummering

Journalen vil gradvis bli et arbeidsverktøy for datamaskinene. Semistrukturerte og strukturerte journaler med standardiserte ord for alle medisinske termer som SNOMED CT vil gjøre det enklere å utvikle autonome systemer. Men det er mye som tyder på at naturlig språkanalyse vil kunne finne opplysninger også i ustrukturert journaltekst. Det er også mulig å lage programmer som gjør journalen mer forståelig for pasientene. Det er viktig at norske helsemyndigheter kartlegger og oppmuntrer utviklingen av naturlig språkanalyseprosjekter i Norge.

Litteratur

1. Bekkelund ASK. *Kunstig intelligens i klinikken: seks trender for fremtidens helsetjeneste*. Oslo: Teknologirådet, 2021. <https://teknologiradet.no/publication/kunstig-intelligens-i-klinikken-seks-trender-for-fremtidens-helsetjeneste/> (14.10.2022).
2. Vallevik VB, Brekke PH. Kunstig intelligens i klinikken – en industri uten råvarer. *Tidsskrift for Den norske legeförening* 2022; <https://doi.org/10.4045/tidsskr.22.0084>
3. *Lov om endringer i helsepersonelloven og pasientjournalloven (bruk av helseopplysninger for å lette samarbeid, læring og bruk av kunstig intelligens i helse- og omsorgstjenesten mv.)* <https://lovdata.no/dokument/LTI/lov/2021-06-11-67> (14.10.2022).
4. *Innsiktsrapport. Behov for data til kunstig intelligens i helsetjenesten*. Rapportnummer IE-1096. Oslo: Direktoratet for e-helse, 2022: 27–8. <https://www.ehelse.no/publikasjoner/behov-for-data-til-kunstig-intelligens-i-helsetjenesten> (14.10.2022).
5. Slaughter L, Oyri K, Fosse E. Evaluation of a Hyperlinked Consumer Health Dictionary for reading EHR notes. *Studies in Health Technology and Informatics* 2011; 169: 38–42. <https://doi.org/10.3233/978-1-60750-806-9-38>
6. Brekke PH, Rama T, Pilán I et al. Synthetic data for annotation and extraction of family history information from clinical text. *Journal of Biomedical Semantics* 2021; 12: 11. <https://doi.org/10.1186/s13326-021-00244-2>
7. Dahl FA, Rama T, Hurlen P et al. Neural classification of Norwegian radiology reports: using NLP to detect findings in CT-scans of children. *BMC Medical Informatics and Decision Making* 2021; 21: 84. <https://doi.org/10.1186/s12911-021-01451-8>
8. Berge G, Granmo O, Tveit T. Combining unsupervised, supervised, and rule-based algorithms for text mining of electronic health records: a clinical decision support system for identifying and classifying allergies of concern for anesthesia during surgery. I: Paspallis N, Raspopoulos M, Barry C et al, red. *Information systems development: advances in methods, tools and management (ISD2017 Proceedings)*. Larnaca, Cyprus: University

- of Central Lancashire Cyprus, 2017. <http://aisel.aisnet.org/isd2014/proceedings2017/CogScience/2> (14.10.2022).
9. Development of a Natural Language Processing (NLP) Web Service for Public Health Use. *Assistant Secretary of Planning and Evaluation, US Department of Health and Human Services*. <https://aspe.hhs.gov/development-natural-language-processing-nlp-web-service-public-health-use> (14.10.2022).
 10. Kreimeyer K, Foster M, Pandeya A et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *Journal of Biomedical Informatics* 2017; 73: 14–29. <https://doi.org/10.1016/j.jbi.2017.07.012>

Erik Fosse
e.t.fosse@medisin.uio.no
Postboks 4950 Nydalen
Oslo universitetssykehus Rikshospitalet
0424 Oslo

Erik Fosse er spesialist i generell kirurgi og i thoraxkirurgi, professor emeritus ved Universitetet i Oslo og spesialrådgiver ved Intervensjonsenteret, Oslo universitetssykehus (www.ivs.no).