

Norsk analyseverktøy

FORSTÅR ALLE SPRÅK

Professor får forsker-
støtte etter lovende
arbeid med
kunstig intelligens

Et nytt norsk produkt skal kunne å forstå og analysere store mengder tekst på et hvilket som helst språk på 48 timer, uavhengig av slang, forkortelser eller skrivefeil.

Språkverktøyet **ANZYZ CCL** er resultatet av 10 års forskning og er basert på kunstig intelligens. Verktøyet skal ha evnen til å lære seg ethvert språk etter å ha blitt foret med store mengder tekst. Det skal også være både selvlærende og selvanalyserende. Mannen bak programmet, professor ved Universitetet i Agder Ole-Christoffer Granmo, forteller til Aftenposten Vitenskap at systemet lærer seg ikke bare å forstå enkeltord, men skjønner hvordan ordene fungerer i sammenhengen og er dermed i stand også til å skjønne både uttrykk og egenskaper såvel som ironi, dobbeltbetydninger og ny-ord...

JO MER, JO BEDRE

Granmo forklarer at Anzyz i hovedsak bruker en metode som språkforskerne kaller korpus-linguistikk sammen med kunstig intelligens. Det innebærer at verktøyet lærer et nytt språk ved at det gjennomgår store mengder med tekst på det aktuelle språket. Jo større mengde tekst, dess bedre språkforståelse. Hvert eneste tegn, ord og uttrykk gjennomgås i detalj. Dette skjer ved å kartlegge hvordan tegnene, ordene og uttrykkene brukes i praksis. På denne måten bygger verktøyet opp språkforståelsen.

Professor Ole-Christoffer Granmo er spesialist på kunstig intelligens og arbeider ved Universitetet i Agder.

FAKTA OM NORSKE ANYZ

- Utviklet ved Universitetet i Agder
- Basert på kunstig intelligens og analyse av store mengder data
- Språkuavhengig. Kan lære et nytt språk på et par dager
- Verktøyet kan identifisere egen- navn
- Kan utføre sentimentanalyse av et stort tekstmateriale (påvise positiv/negativ omtale)
- Dele opp en stor tekst i tema,

- Kjenne igjen beslektet materiale
- Kartlegge relasjoner og sammenhenger, f.eks finne sammenheng mellom symptomer og sykdommer i et stort tekstmateriale
- Utfører foreløpig ikke maskinoversettelse
- Har mottatt Sparebankens Sørs og Diggins Innovasjonspris for 2015
- Har mottatt støtte fra Forskningsrådets program Forny2020

FORSKNINGSRÅDET: - SPENNENDE PRODUKT

Wibke Ramm i avdeling for forskningsdrevet innovasjon ved Norges Forskningsråd bekrefter at Anzyz-prosjektet har fått innvilget støtte på kr 3,845 millioner kroner fra programmet «FORNY2020», for en prosjektperiode på litt under to år, fra 2015 til 2017. FORNY2020 er Forskningsrådets program for å bringe resultater fra offentlig finansierte forskningsinstitusjoner fram til markedet. Midlene skal bidra til å avklare spørsmål knyttet til kommersielle, forretningsmessige og teknologiske aspekter.



FOTO © UIA

Det nye språkverktøyet som Granmo nå arbeider for å kommersialisere er resultatet av 10 års forskning og utvikling.

med potensielle kunder som Telenor og Sørlandet sykehus ble i følge henne positivt vurdert og prosjektet fikk en høy total karakter i fagpanelets vurderinger. – Denne vurderingen er grunnlaget for støtten til prosjektet. Forskningsrådet bidrar med opp til 80% av kostnadene i verifiseringsprosjekter. Dette er et spennende produkt, sier hun

– Når mønstrene i språket er kartlagt, gir man eksempler for å guide systemet. Hvis brukeren eksempelvis ønsker å analysere finansprat på språket ukrainsk, gir han systemet noen få eksempler på relevante begreper, forklarer han

I løpet av brøkdelen av et sekund kommer Anzyz-systemet tilbake med et forslag som brukeren gir feedback på. Deretter er analyseoppsettet klart. Målet er i følge UiA-professoren å få et fruktbart samarbeid mellom menneske og kunstig intelligens.

Programmet skal i følge UiA-professoren være lett å bruke. – Man kan utføre avansert tekstanalyse på en veldig brukervennlig og effektiv måte, sier han.

Anzyz er likevel ikke et verktøy som

kan sammenlignes med Google Translate. Det kan nemlig foreløpig ikke brukes til maskinoversettelser. Derimot har det flere andre bruksmuligheter. Granmo er i full gang med kommersialisere verktøyet gjennom sitt eget firma, Anzyz Technologies AS.

I første omgang arbeider han med å undersøke bruksmulighetene til språkverktøyet på tre hovedområder, til kundebehandling i store bedrifter, innenfor helse og medisin og innenfor sikkerhet og overvåkning.

GAL MASKINOVERSETTELSE

Telenor er blant storselskapene som har testet programmet fra Anzyz. Forskningslederen Jarle Hildrum har store forventninger: – For å øke presi-

sjonen på analysene på mindre språk, er det viktig at systemet kan læres opp til å tolke det aktuelle språket, sier Hildrum.

Seniorforsker samme sted, Knut Kvale, forteller at Telenor mottar flere hundre tusen skriftlige tilbakemeldinger fra kunder hvert år. Selskapet bruker tilbakemeldingene for å iverksette forbedringer. Hver enkelt melding blir gjennomgått og kategorisert ut i fra hva det gjelder. – Denne første kategoriseringen ønsker vi å effektivisere med automatisk tekstanalyse, sier Telenors seniorforsker.

Ettersom telegiganten opererer i 13 land, trenger selskapet tekstanalyse som fungerer for tilsvarende antall språk. Det eksisterer i følge Kvale flere slike analyseverktøy for engelsk, men

disse forstår ikke alle språkene i landene Telenor opererer i, som for eksempel norsk eller bengali. Nåværende rutiner innebærer at systemene først må oversette teksten til engelsk med et automatisk oversettelsesprogram, som Google Translate, og deretter tolke og kategorisere den engelske teksten.

– Dette bærer noen ganger helt galt av sted fordi oversettelsesprogrammet gjør feil, sier Kvale.

KAN REDDE LIV

Sørlandet Sykehus tester også analyseverktøyet Anzyz. Doktorgradsstipendiat Geir Thore Berge og forskningssjef Frode Gallefoss forsker blant annet på bruk av denne type teknologi i medisinsk sammenheng.

Berge forteller at teknologier som muliggjør automatisk strukturering av fritekstinformasjon i pasientjournal og avansert klinisk beslutningsstøtte nå er på vei inn i helsevesenet etter å ha vært brukt i andre sammenhenger en stund.

Helse-Norge har i følge ham enorme mengder med fritekst-basert informasjon i sine pasientsystemer.

– Frem til nå har det vært vanskelig å bruke denne fritekstinformasjonen effektivt i medisinsk sammenheng, forteller han.

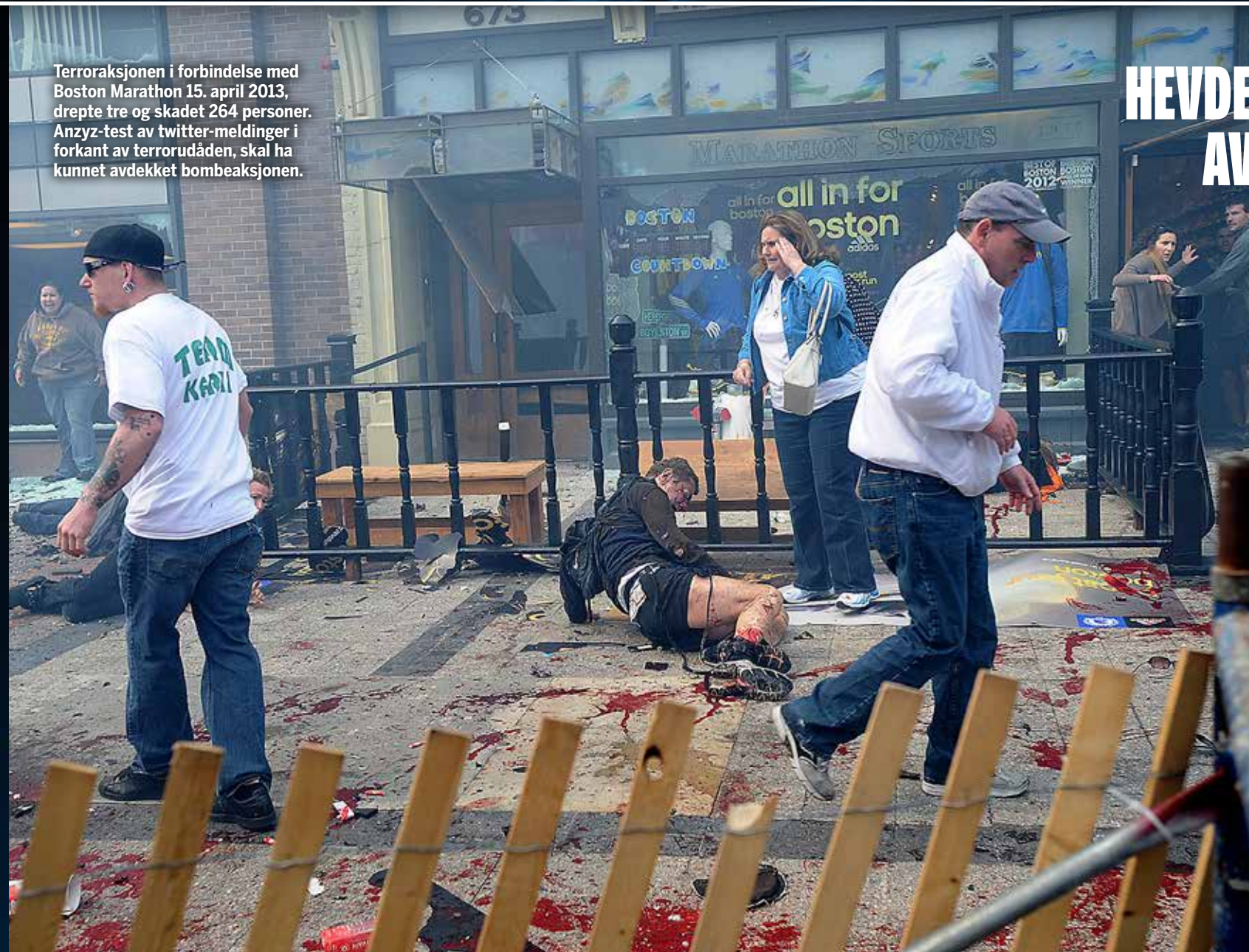
Det dreier seg om store mengder medisinske data som er pakket inn i mange ulike formater, som skannede journaldokumenter, blanketter, skjemaer og pdf-dokumenter som ikke kan gjennomføres på samme måte

som strukturerte data. Slik dokumentasjon kan i følge ham i en del tilfeller inneholde informasjon som kan ha livsviktig betydning for fremtidig pasientbehandling.

– I verste fall kan slik informasjon drukne i en stor mengde annen informasjon, og dermed ikke nå frem til helsepersonell som skal behandle pasienter, sier han.

Agder-forskeren skal nå trekke ut fritekstdata fra pasientsystemet ved Sørlandet Sykehus som deretter skal videreanalyseres med det nye språkverktøyet Anzyz.

AV STEINAR SUND



Terroraksjonen i forbindelse med Boston Marathon 15. april 2013, drepte tre og skadet 264 personer. Anzyz-test av twitter-meldinger i forkant av terrorudåden, skal ha kunnet avdekket bombeaksjonen.

HEVDER SPRÅKVERKTØYET KUNNE AVDEKKET BOSTON-TERROREN

Professor Ole-Christoffer Granmo og hans medarbeidere prøver nå å finne ut hvilke muligheter språkssystemet gir innenfor sikkerhet og forebygging av terror. Anzyz-verktøyet skal nemlig i følge ham også kunne arbeide i sanntid og lese gjennom store mengder med meldinger på sosiale nettsteder etter hvert som de blir lagt ut.

Programmet skal kunne lære seg hva som er normal nettpat og skal kunne reagere når det skjer noe uvanlig. Eksempelvis skal programmet ha potensial til å reagere hvis det oppdager våpenprat eller diskusjoner om kjøp av våpen når det foreligger koblinger til visse ideologier.

– Den lærer hvordan denne type grupperinger snakker sammen og kjenner det igjen, forklarer Granmo.

Systemet skal også kunne påvise det hvis det er spesielt mye hets rundt en bestemt offentlig person slik at det skal kunne gi innspill til risikovurderinger ved å analysere meldingsstrømmene på sosiale medier.

En test viste i følge Granmo at programmet kunne ha vars-



Ole-Christoffer Granmo

FOTO © UIA

let om Boston-bomberne på et tidlig tidspunkt ved hjelp av analyse av Twitter-meldinger om det hadde vært brukt. Målet er å kunne avdekke og stoppe hendelser i forkant.

KARTLEGGE IS-REKRUTTERING

Nå gir programmet muligheter til å følge med på rekrutteringen til IS.

– IS rekrutterer jo nettopp sine folk på sosiale medier. Programmet kan fange opp og kartlegge denne typen rekrutteringer, sier Granmo.

Om UiA-professoren har rett, kan dette verktøyet dermed gi uvurderlig hjelp til alle som har tillatelse til å arbeide med sikkerhetsovervåkning.

