



(12) **PATENT**

(19) NO

(11) **330017**

(13) **B1**

NORGE

(51) Int Cl.

G06F 7/00 (2006.01)

G06F 17/21 (2006.01)

G06F 17/30 (2006.01)

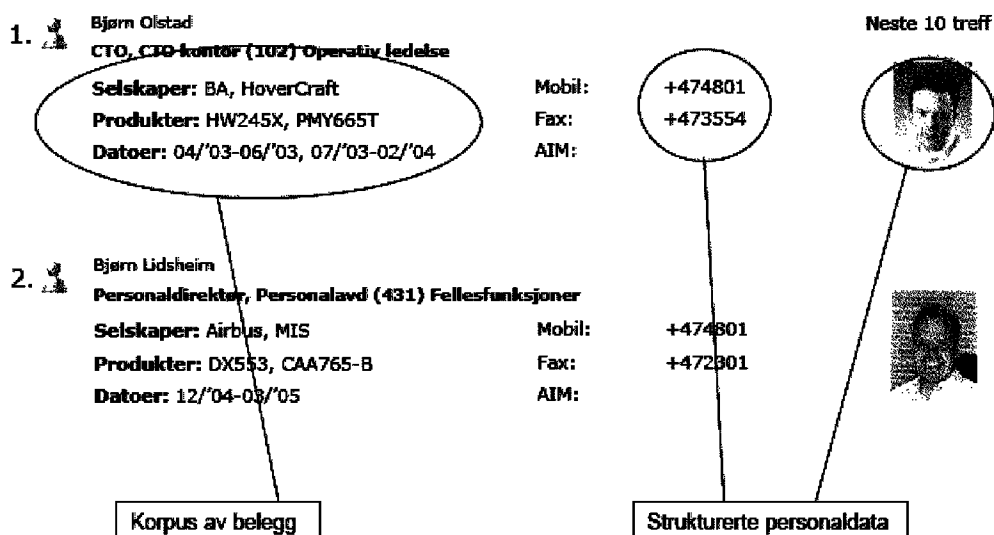
G06N 5/00 (2006.01)

Patentstyret

(21)	Søknadsnr	20091341	(86)	Int.inng.dag og søknadsnr
(22)	Inng.dag	2009.03.31	(85)	Videreføringsdag
(24)	Løpedag	2009.03.31	(30)	Prioritet
(41)	Alm.tilgj	2010.10.01		
(45)	Meddelt	2011.02.07		
(73)	Innehaver	Microsoft International Holdings B.V., Telestone 8 - Teleport, Naritaweg 165, NL-1043BW AMSTERDAM, Nederland		
(72)	Oppfinner	Damien Islam-Frénoy, 18 Warren Close, GB-HP178YL STONE BUCKS, Storbritannia		
(74)	Fullmektig	Bryn Aarflot AS, Postboks 449 Sentrum, 0104 OSLO, Norge		

(54)	Benevnelse	Kriteriebasert lokalisering av kontekstuellt relevante størrelser
(56)	Anførte publikasjoner	US2004/243556 A1, WO2005/074410 A2
(57)	Sammendrag	

I en fremgangsmåte for å lokalisere ekspertstørrelser forbundet med en bestemt kontekst i dokumenter og innehold indeksert og gjort søkbart på en søkemotor, dannes logiske regler for å identifisere størrelsen uttrykt ved ekspertise i et dokument, et dokumentbehandlingssystem benyttes på identifikasjonen, og de behandlede dokumenter indekseres i en søkemotor. En navigator settes opp i søkemotoren for ekspertområder, og et søkespørsmål gis til søkemotoren for å gjenfinne en resultatmengde av ekspertstørrelser. Søkeresultatet vises for en bruker, med hvert treff i søkeresultatet som en ekspertstørrelse og forbundet informasjon. Fremgangsmåten kan benyttes både på individsentrerte ekspertstørrelser så vel som kollektivsentrerte ekspertstørrelser som inneholder ekspertstørrelser i form av steder med høy verdi.



INNLEDNING

Den foreliggende oppfinnelse angår en fremgangsmåte for å lokalisere ekspertstørrelser relatert til en bestemt kontekst i dokumenter eller innhold indeksert og gjort søkbart på en søkemotor.

5

Spesielt tillater den foreliggende oppfinnelse innføring av slike størrelser fra et korpus av dokumenter basert på valgte kriterier (konteksten). For et bestemt emne eller kontekst kan en slik størrelse identifiseres som en ekspert eller høyverdig node. Hvor det ikke er tilstrekkelig informasjonskilder som teller opp og beskriver den totale mengde av størrelser, kan likevel fremgangsmåten i henhold til den foreliggende oppfinnelse gi meget nøyaktige resultater, f.eks. ved å se etter folk forbundet til konteksten, hvor ingen autoritativ liste over personer og en beskrivende lister over deres interesseområder og ekspertise foreligger. Fremgangsmåten i henhold til den foreliggende oppfinnelse er søkebasert og kan således benyttes på forbehandlet og søkbart innhold og dokumenter.

15

KJENT TEKNIKK

Nedenfor gis en oversikt over fremgangsmåter kjent i teknikken. Noen av disse fremgangsmåtene er basert på å danne en statisk dokumentdatapost eller profil for størrelsen, mens andre forsøker å oppdatere størrelsespostene etter hvert som nye dokumenter publiseres eller innhold forbrukes.

20

Dannelse av statiske poster ville gjøre det nødvendig å telle opp hele universet av størrelser og deretter benytte en manuell tilnærming for å frembringe en statisk deskriptor for hver enkelt. Når det søkes etter relevante størrelser i kontekst, ville disse data bli behandlet (f.eks. på en søkemotor) for å generere en liste av størrelser hvis deskriptorer tilsvarer konteksten. Dette er iboende altfor ofte overfladisk og blir hurtig utdatert.

25

Dynamisk oppdatering av dataposter er lik den ovennevnte tilnærming, idet separate dataposter settes opp for hver størrelse. I dette tilfelle kan forbundne innholdseventer (så som deteksjon av et nytt dokument som nevner størrelsen) trigge en oppdatering av hele deskriptorlageret. Dette vil fortsatt danne en enkelt de-

30

skriptor pr. størrelse, og vil følgelig iboende la være å ta hensyn til noe av konteksten til de individuelle dokumenter (ytterligere størrelser eller datoer) så vel som sikkerhetsinformasjon som kan begrense forbruket av innhold.

5 I en tredje metode kjent i teknikken benyttes en søkemotordrevet navigator som lister forekomstfrekvensen av størrelsen innenfor en kontekstuell undermengde av dataene, til å belyse de mest relevante størrelser. Dette avhenger av at størrelsen eksisterer eller er blitt identifisert som et metadatafelt i dokumentkorpuser. Denne navigatormetoden overviner de statiske fallgruver ved de første to metoder, men
10 på grunn av et altfor forenklet relevanstilnærming fører den typisk til resultater med dårlig kvalitet.

De fra kjent teknikk nevnte metoder lider således alle av ikke å være i stand til å frembringe en kvalitetslokkemiddel eller sammendrag som forklarer søkeren hvorfor
15 for et bestemt ekspertstørrelse angår et emne.

OPPFINNELSENS HENSIKTER

Følgelig er det en hovedhensikt med den foreliggende oppfinnelse å lokalisere og identifisere ekspertstørrelser som forekommer i dokumenter og forbinde emner
20 (konteksten) med de størrelser som de har størst relevans for.

En annen hensikt med oppfinnelse er å tillate relevant størrelsesidentifikasjon over en rekke kilder. Disse relevante størrelser er kjent som "eksperter".

25 Nok en annen hensikt er dynamisk å generere beleggmengde for å validere forbindelsen mellom kontekst og størrelser.

En siste hensikt med oppfinnelsen er at lokaliseringen og identifikasjonen skal utføres ved å benytte en søkemotor.

30

SAMMENDRAG AV OPFFINNELSEN

De ovennevnte hensikter så vel som ytterligere trekk og fordeler realiseres med en fremgangsmåte i henhold til den foreliggende oppfinnelse som er kjennetegnet

ved å omfatte trinn for å danne logiske regler for å identifisere størrelser uttrykt ved ekspertise i et dokument, å anvende et dokumentbehandlingssystem som bruker de logiske regler for å foreta slike identifikasjoner i dokumentene og å indeksere således behandlede dokumenter i søkemotoren, å etablere en navigator i søkemotoren for i det minste noen områder av ekspertise, å inngi et søkespørsmål til søkemotoren for utvinne en resultatmengde i form av en sortert liste av ekspertstørrelser på basis av navigatoren for ekspertiseområder, og å vise søkeresultatet for en bruker, idet hvert treff i søkeresultatet utgjør en spesifikk ekspertstørrelse og forbunden informasjon.

Fremgangsmåten i henhold til oppfinnelsen skal nå beskrives mer detaljert nedenfor, med henvisning til den vedføyde

figur 1 som viser en resultatmengde av ekspertstørrelser gjenfunnet med fremgangsmåten i henhold til den foreliggende oppfinnelse.

DETALJERT DRØFTELSE

Fremgangsmåten i henhold til den foreliggende oppfinnelse muliggjør identifikasjon og lokalisering av ekspertstørrelser, dvs. i de mest relevante størrelser, gitt en eller flere informasjonskilder. En ekspertstørrelse er definert som den mest relevante for et bestemt emne.

Ekspertlokaliseringer som benytter høykvalitetsnormalisering, dynamisk relevantbasert ekspertidentifikasjon og oppdagelse av bekreftende belegg er en nøyaktig og lett anvendelig metode for å identifisere relevante størrelser med hensyn til et bestemt emne og kan appliseres over en rekke innholdsinformasjonskilder. For hvert av dataelement i en kilde, benyttes dataekstraksjon og rensing for å fremheve i kontekst en eller flere relaterte størrelser innenfor hvilke typer hvilke eksperter vil finnes (type A størrelser) så vel som andre størrelsestyper (type B størrelser). Dataene indikseres i en søkemotor som støtter rangavstemning og navigasjon (med glidende relevans). Søk som innbefatter hvilken som helst sikkerhets- eller filtreringskrav vil returnere en liste av relaterte type A-størrelser. Sekundære søkespørsmål benyttes til å identifisere den samlede kontekst som lenker hver

størrelse til emnet. Denne informasjon fremlegges som dynamiske lokkemidler og gjør følgelig brukeren i stand til å forstå resultatene.

5 Det første trinn i fremgangsmåten i henhold til foreliggende oppfinnelser er å danne logikk for å identifisere hvordan type A-størrelser relatert til hvert dokument defineres for hver kilde eller dokument. For eksempel dersom type A var personer, kunne dette være forfatterskap eller referanse til en bestemt kontekst på siden. Regler blir dannet innenfor en dokumentprosesserende ramme for å markere disse i konteksten. Hvis noen størrelser har en sterkere forbindelse til dokumentet, 10 kan de også bli vektet henholdsvis. I tillegg kan datarenseteknikker benyttes til å normalisere forskjellige varianter av samme størrelse (dvs. for personen Tony Hart, Dr. Anthony Heart...). Lignende teknikker benyttes til å identifisere type B-størrelser (f.eks. bedrifter, steder eller industrispesifikke begreper).

15 Hvert dokument i systemet fra alle relevante kilder går gjennom de ovennevnte opplegg for dokumentbehandling. Følgelig blir dokumenter markert med forskjellige relaterte type A- og type B-størrelser spesifisert i normalisert form så vel som andre strukturerte metadata. Disse blir deretter indeksert i en søkemotor som har et navigasjonsoppsett for disse feltene.

20 Slik det skal forstås er type A en ekspertstørrelse gitt som folk eller personer, og det skal således forstås at type A-størrelser er identifisert med individer, selv om dette ikke betyr at individene fysisk skal identifiseres med personer. De kunne også identifiseres med enhver selvopprettet høyverdig sted og således en bestemt innholdskilde, sted eller dokument. Ekspertstørrelser av type B skiller seg 25 ved at de høyverdige steder kan identifiseres som understørrelser inneholdt innenfor type B-størrelser, med andre ord, interne steder for kunnskap i et foretak eller organisasjon. Fysisk kan de naturligvis i seg selv identifiseres med personer innenfor denne bestemte ramme. En søkemotor på hvilken fremgangsmåten i henhold til oppfinnelsen er implementert, kan således omfatte en navigator som er 30 dannet for størrelsesklassifisering på forskjellige nivåer, og deretter kan disse passende navigators integreres med navigasjonsiverksatte felter for søking etter ekspertise. Dette kan være spesielt verdifullt når slike høyverdige, interne steder for

kunnskap er rommet innenfor større ramme som omfatter et komplekst nettverk av steder som ikke nødvendigvis har noen relevans for søking og lokalisering av ekspertstørrelser. Dette vil mer ofte enn ikke være tilfelle for organisasjoner og foretak. Det vil følgelig være svært nyttig for ansatte og andre operatører innenfor en organisatorisk ramme å ha til rådighet mest mer effektive måte for å lokalisere og gjenfinne relevant kunnskap innenfor denne rammen.

For å finne en ekspert i et område Q blir et søkespørsmål for Q i dokumentinnholdet sendt til søkemotoren. Søkespørsmålet som legges inn kan inneholde strukturerte filtrerte kriterier så som adgangsrestriksjoner. Søkemotoren eksekverer søket for $Q \cap F$ og returnerer en sortert liste av type A størrelser basert på ekspertfeltnavigatoren definert ovenfor. Sorteringsmetoden innbefatter, men er ikke begrenset til

- Forekomstfrekvens i søkeresultatmengden
- Forekomstfrekvens i de øverste dokumenter M hvor dokumentene sorteres av en relevansalgoritme,
- Frekvens vektet etter dokumentrelevans

Gitt at ekspertlokaliseringsgrensesnittet viser N resultater, velges de øverste N størrelser i denne sorterte liste. For hvert element P i denne liste blir et søkespørsmål gitt til indeksen for $Q \cap F \cap P$, dvs. et søkespørsmål for dokumenter som svarer til søkespørsmålet i et opprinnelig filter eller et filter på ekspertattributten. For dette søkespørsmål vil motoren returnere en eller mange navigatører basert på andre filtre (f.eks. de ekstrahert under dokumentbehandlingstrinnet). Sorteringsalgoritmens muligheter for disse navigatører er som ovenfor. Denne informasjonen er det korpus av evidens som viser størrelsens P s forbindelse med Q . Den benyttes til å danne en dynamisk lokkemiddel for resultatet (lokkemiddelet for P_n er størrelsene fra $Q \cap F \cap P_n$).

Samlet blir det derfor sendt $N+1$ søkespørsmål (opprinnelig søkespørsmål og N for korpus av belegg). Ytterligere kan det gis en anmodning om nyttig virkningsiverksettende informasjon fra passende kilder eller indeks.

Resultatene blir deretter vist for brukeren, idet hvert treff er en ekspert med sin deskriptor, dynamiske lokkemiddel og ytterligere informasjon. Et eksempel på en resultatmengde for eksperter i form av personer er vist på fig. 1, hvor $N=2$ type A entiteter, eller eksperter, på et bestemt ekspertiseområde er vist på høyre side. I eksempelet på fig. 1 er hver ekspert listet med et portrett, mobiltelefonnummer osv, som et resultat av en anmodning om nyttig virkningsiverksettende informasjon fra en strukturert personalkilde i dette bestemte eksempelet. For hver av ekspertene P_1 , P_2 er det utført et søkespørsmål for å fremskaffe et dynamisk lokkemiddel eller korpus av belegg. Resultatene er vist på venstre side, og inkluderer navn og ekspertiseområde samt type B entiteter. I eksempelet på utførelsesform på figur 1 er type B entitetene vist som Selskaper, Produkter og Datoer.

Fremgangsmåten i henhold til den foreliggende oppfinnelse skaffer en rekke fordeler med hensyn til å få informasjon om hvor en bestemt ekspertise er lokalisert, enten den er basert på metadata eller satt inn i en bestemt kontekst eller ramme innenfor en dokumentkilde. I tilfelle menneskelig (personer) eksperter er fremgangsmåten i henhold til den foreliggende oppfinnelse spesielt anvendbar for eksempel i forbindelse med kunnskapsforvaltning eller samarbeidssystemer for å identifisere innvendige kunnskapssteder i et foretak eller organisasjon. Den kan også benyttes til vevsteder med blogger eller nyheter for å identifisere opinionsdannere. Andre applikasjoner vil f.eks. være informasjonsskanker, meldingstavler eller andre steder hvor individer deler kunnskap.

PATENTKRAV

1. Fremgangsmåte for å lokalisere ekspertstørrelser forbundet med en bestemt kontekst i dokumenter og innhold indeksert og gjort søkbart på en søkemotor, hvor
5 fremgangsmåten omfatter trinn for
- a) å danne logiske regler for å identifisere størrelser uttrykt ved ekspertise i et dokument,
 - b) å anvende et dokumentbehandlingssystem som bruker de logiske regler for å foreta slike identifikasjoner i dokumentene og å indeksere således behandlede
10 dokumenter i søkemotoren,
 - c) å etablere en navigator i søkemotoren for i det minste noen områder av ekspertise,
 - d) å inngi et søkespørsmål til søkemotoren for utvinne en resultatmengde i form av en sortert liste av ekspertstørrelser på basis av navigatoren for ekspertiseområder,
15 og
 - e) å vise søkeresultatet for en bruker, idet hvert treff i søkeresultatet utgjør en spesifikk ekspertstørrelse og forbunden informasjon.
2. Fremgangsmåte i henhold til krav 1,
20 k a r a k t e r i s e r t v e d å bestemme en virkelig relasjon mellom ekspertstørrelse og et dokument eller en innholdskilde.
3. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å benytte datarensing for å normalisere forskjellige
25 varianter av de samme størrelser.
4. Fremgangsmåte i henhold til krav 3,
k a r a k t e r i s e r t v e d å spesifisere ekspertstørrelsen som forbundet med et gjenfunnet dokument i en normalisert form for en størrelsesmerking i dokumentet.
30
5. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å klassifisere ekspertstørrelser som sentrert på indivi-
der eller kollektiver.

6. Fremgangsmåte i henhold til krav 5,
k a r a k t e r i s e r t v e d ytterligere å underklassifisere de individsentrerte eks-
p e r t s t ø r r e l s e r i henhold til et fysisk klassifikasjonsskjema og å danne en navigator
5 for å ekstrahere spesifikke underklasser av ekspertstørrelser, idet underklassene i
seg selv er oppført under en definert ekspertiseområde.

7. Fremgangsmåte i henhold til krav 5,
k a r a k t e r i s e r t v e d å identifisere de kollektivsentrerte ekspertstørrelser
10 som ekspertstørrelser som inneholder en eller flere steder med høy verdi og for-
bundet med et definert ekspertiseområde.

8. Fremgangsmåte i henhold til krav 5,
k a r a k t e r i s e r t v e d å underklassifisere en eller flere steder i henhold til et
15 fysisk klassifikasjonsopplegg, og å sette opp en navigator for å ekstrahere et be-
stemt sted som en ekspertstørrelse i seg selv.

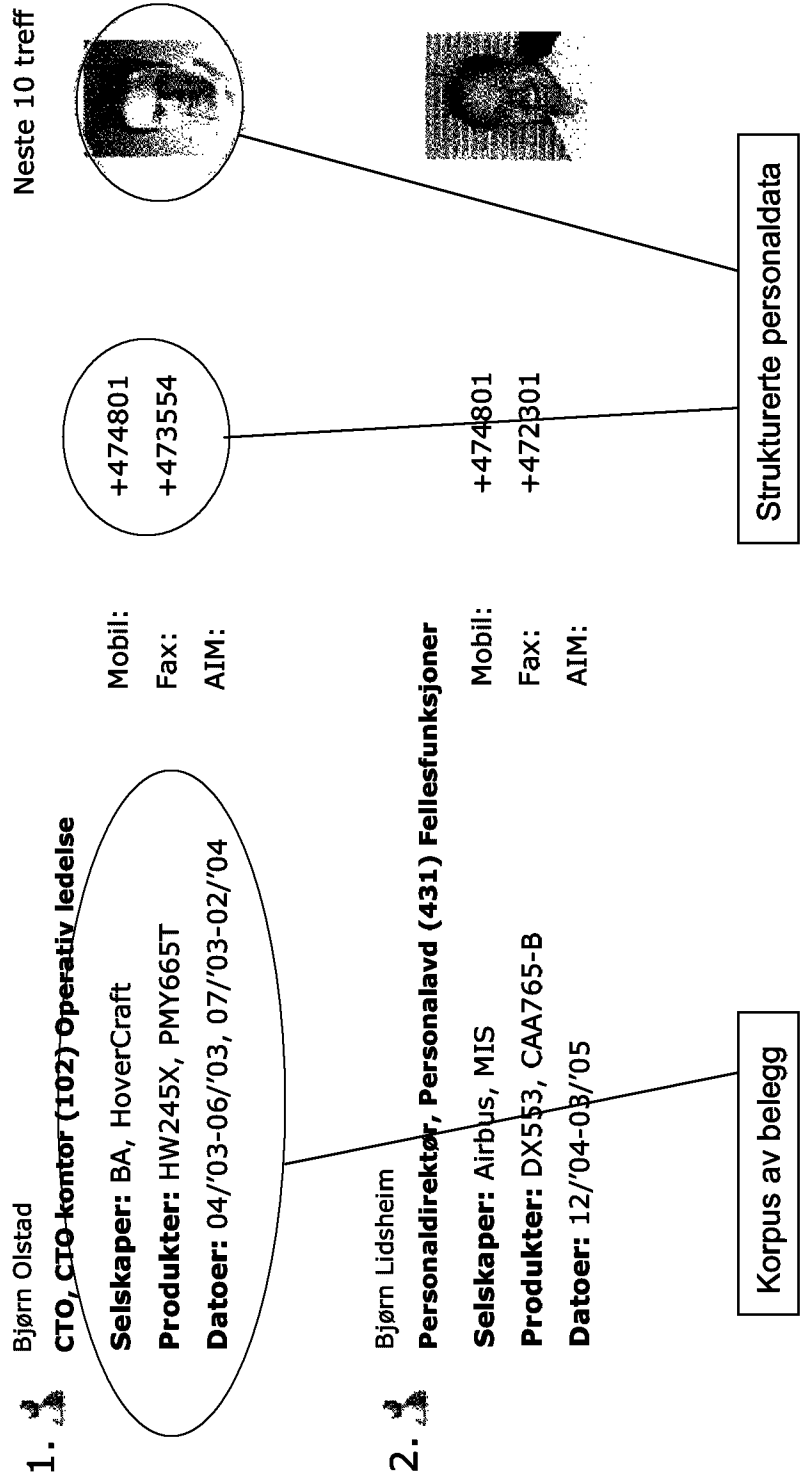


Fig. 1