(12) **PATENT**

(19) NO (11) **326743** (13) **B1**

NORGE (51) Int Cl.
G06F 17/30 (2006.01)

Patentstyret

(21)	Søknadsnr	20075351	(86)	Int.inng.dag og søknadsnr
(22)	Inng.dag	2007.10.18	(85)	Videreføringsdag
(24)	Løpedag	2007.10.18	(30)	Prioritet
(41)	Alm.tilgj	2009.02.09		
(45)	Meddelt	2009.02.09		

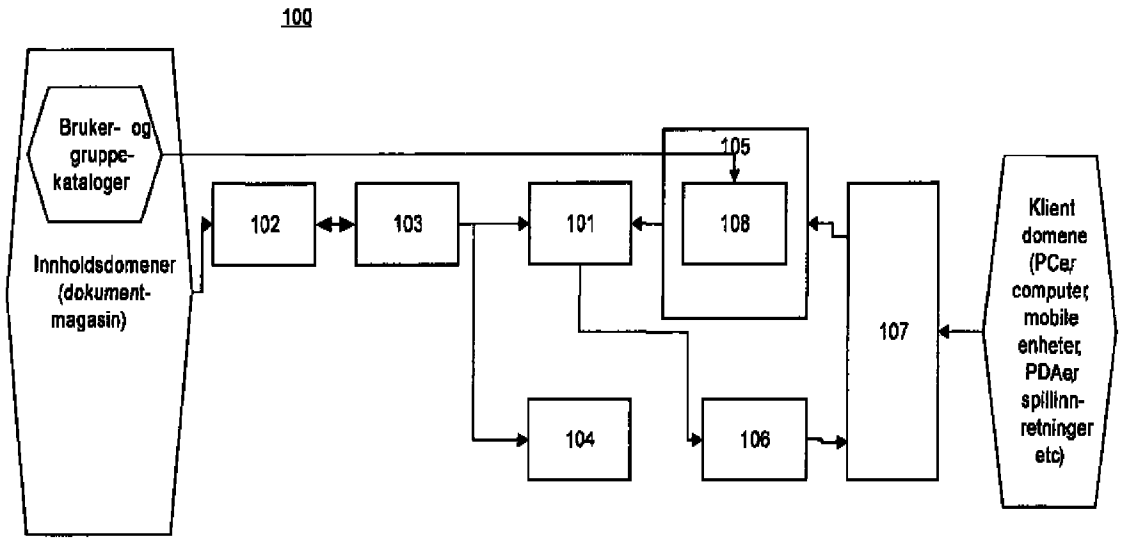
(73) Innehaver Fast Search & Transfer ASA, Postboks 1677 Vika, 0120 OSLO
(72) Oppfinner Anund Lie, Åsdalsveien 52 A, 1167 Oslo
Helge Grenager Solheim, Hellerudfaret 8b B, 0672 OSLO
Øystein Hallaråker, Observatoriegt. 12, 0254 OSLO
(74) Fullmektig

(54) Benevnelse **Fremgangsmåte for å begrense aksess til søkeresultater og søkemotor som støtter fremgangsmåten**

(56) Anførte publikasjoner US 2006/0020587, US 2007/0067270, US 7,031,954

(57) Sammendrag

I en fremgangsmåte for aksess, søking og gjenfinning av informasjon over et datakommunikasjonssystem generelt, hvor et søkespørsmål benyttes på et sett av dokumenter, blir en resultatmengde fra de tilsvarende dokumenter identifisert. Fremgangsmåten omfatter å revidere søkespørsmålet i henhold til den aktuelle brukers aksesstillatelser til originaldokumentene i et kildesystem på en slik måte at bare dokumenter som brukeren tillater å aksessere direkte fra forskjellig kildesystemer, forekommer i resultatmengden, selv når kildedokumentene befinner seg i systemer med forskjellige sikkerhetsdomener som potensielt er avhengige av hverandre. I en søkemotor (100) som er i stand til å støtte og implementere den ovennevnte fremgangsmåte, omfatter søkemotoren som i og for seg kjent undersystemer for å utføre søking og gjenfinning i form av én eller flere kjernesøkemotorer (101), et applikasjonsprogrammert grensesnitt (102) for innhold, et innholdsanalysetrinn (1 03) og et applikasjonsprogrammert grensesnitt (107) for en klient forbundet med kjernesøkemotoren (101) via analysetrinn (105;106) for søkespørsmål og søkeresultat. I tillegg omfatter søkemotoren (100) for å støtte den ovennevnte fremgangsmåte en modul (108) for å revidere søkespørsmålet.



Den foreliggende oppfinnelse angår en fremgangsmåte for å begrense aksess til søkeresultater i form av dokumenter hentet fra et dokumentmagasin, hvor fremgangsmåten for å begrense aksess til søkeresultater i form av dokumenter hentet fra et dokumentmagasin, hvor fremgangsmåten angår et system for aksessering eller søking av informasjon, hvor en bruker av systemet for søking eller aksessering benytter et søkespørsmål på dokumentmagasinet for å innhente en resultatmengde i form av dokumenter fra dette, hvor aksessen er begrenset til de dokumenter i resultatmengden eller alle gjenfunne dokumenter som har en aksesskontrolliste som tilsvarer et filter utført som et søkespørsmål, hvor systemet for aksess eller søking av informasjon er implementert på en søkemotor.

Den foreliggende oppfinnelse angår også en søkemotor for å støtte og implementere fremgangsmåten i systemer for aksess eller søking av informasjon hvor søkemotoren benyttes til å aksessere, søke, gjenfinne og analysere informasjon fra dokument- eller innholdsmagasiner som er tilgjengelige via datakommunikasjonsnettverk, innbefattet ekstranett og intranett, og å fremlegge søke- og analyseresultater for sluttbrukere, hvor søkemotoren omfatter minst en kjernesøkemotor, et applikasjonsprogrammert grensesnitt for innhold (innholds-API) forbundet med den minst ene kjernesøkemotor via et innholdsanalysetrinn, og et applikasjonsprogrammert grensesnitt for søkespørsmål forbundet med den minst ene kjernesøkemotor via respektive søkespørsmålsanalyse- og resultatanalysetrinn.

Informasjonsgjenfinning har tradisjonelt beskjeftiget seg med å indeksere data fra flere kilder. Aksesskontroll til dokumentene er blitt løst ved etterfiltrering av resultatmengdene med bruk av oppkall fra applikasjonsprogrammerte grensesnitt (API) mot hvert kildesystem. Dette har alvorlige konsekvenser for søkelatensitet og gjør effektive dype navigatorer umulige i praksis. Alternativt har søkeindeksen blitt satt i stand til å indeksere aksesskontrollinnførsler med dokumentene ved å etterligne aksesskontrollmekanismen til kildesystemene, og søkespørsmålene er blitt omskrevet i henhold til brukerens aksessadganger. For denne løsning har bare dokumenter fra kompatible sikkerhetsdomener blitt tillatt i resultatmengdene. Noen ganger har begrensede identitetsavbildningsmekanismer blitt benyttet for i en viss grad å støtte forskjellige sikkerhetsdomener.

I det følgende er begrepet "dokument" benyttet for ethvert søkbart objekt og det kunne for eksempel bety et tekstdokument, et dokument representert i XML, HTML, SGML, eller et kontorformat, et databaseobjekt så som post, tabell, visning eller søkespørsmål, eller et multimediaobjekt. Følgelig skal "dokument" bli betraktet som synonymt med "innhold".

Aksesstillatelsene til en bruker som aksesserer et informasjonssystem, blir bestemt av mengden av grupper som brukeren er medlem av. Brukerne kan være direkte eller indirekte medlemmer av gruppene ved å være medlemmer av grupper som i seg selv er medlemmer i andre grupper. For således å finne den samlede mengde av grupper er det nødvendig å utføre en uttømmende gjennomløping av denne medlemskapsgrafen, noe som vil være meget tidkrevende når det er et stort antall brukere og grupper i sikkerhetsdomenet. Når en aksesskontroll imidlertid benyttes konvensjonelt, blir medlemskapene evaluert bare for et enkelt domene. Den ovennevnte etterfiltrering av søkeresultater er et eksempel på dette.

Fra kjent teknikk foreligger det flere tilnærminger for å forbedre hastigheten til grafgjennomløpingen som er nødvendig for å bestemme gruppemedlemskapene for en gitt bruker. De fleste angår enkeltdomenetilfellet hvor hensikten er å bestemme gruppemedlemskap som bestemmer aksessadganger for en enkelt bruker i et enkelt domene (eller til og med til et enkelt objekt) og som ikke lett lar seg skalere til flerdomenetilfellet, noe som er nødvendig for søking med forfiltergenerering.

For eksempel viser US patent nr. 7 103 784 hvordan grupper kategoriseres som lokale, universelle og globale, og restriksjoner pålegges på hvordan disse kategoriene av grupper kan nøstes. Virkningen er at bare en (antatt liten) undermengde av grupper tas i betraktning for kryssdomenemedlemskap. For grupper med potensielle kryssdomenemedlemskap er det fortsatt nødvendig å konsultere alle domener for å finne ytterligere medlemmer.

US patent nr. 7 085 834 viser en prosess for å bestemme mengden av grupper som brukeren er medlem av, men henvender seg ikke spesifikt til flerdomenetilfelle og har ingen midler for å optimere den rekursive grafgjennomløping som er nødvendig for å oppløse nøstede grupper.

Videre angår US patent nr. 7 076 795 gruppebasert autorisasjon, men viser en spesiell måte å organisere tabellene som avbilder brukeridentifikasjoner til grupper og aksessrettigheter. Det er ikke noen muligheter for nøstede grupper, da den implisitte antagelse er at lukningen av medlemskapsrelasjonen er beregnet på forhånd. Dette skalerer ikke godt når gruppemedlemskapene er dynamiske eller opprettholdes over flere domener.

Endelig angår US patent nr. 7 031 954 en fremgangsmåte og et system for dokumentgjenfinning i et nettverkmiljø med vevtjenere hvor dokumentene er lagret med forskjellige aksessnivåer og hvor søkespørsmål inngis fra søketjenere. Spesifikt angår US patent nr. 7 031 954 etterfiltrering av søkeresultatene. En person som foretar søket, skal ha en entydig identifikasjonskode som imidlertid ikke tar hensyn til aksesskontrollbegrensninger. URLene til dokumentene som returneres i et søk, blir gjennomgått etter at søket er utført, og det benyttes en aksesskontrolliste knyttet til hver dokumenttjener for å kontrollere hvorvidt den angjeldende URL er kompatibel med aksessnivået for identifikasjonskoden til den personen som utførte søket. Bare dokumenter eller nettadresser som er kompatible med søkerens aksessnivå, blir levert ut, mens URLer som ikke er kompatible med personens aksessnivå, holdes tilbake, og vedkommende vil heller ikke få kunnskap om hvilke URLer som ikke er kompatible med angjeldende aksessnivå.

I lys av manglene ved den ovennevnte kjente teknikk er det derfor en første hovedhensikt med den foreliggende oppfinnelse å beskytte dokumenter mot uautorisert aksess, samtidig som det fortsatt skaffes aksess til alle dokumenter som den foreliggende bruker har aksess til i kildesystemet.

En sekundær hensikt med den foreliggende oppfinnelse er å unngå å utføre kostbar etterfiltrering og å konsultere hvert kildesystem som foreligger i resultatmengden som en del av hver syklus for søkespørsmål og svar.

En annen hensikt med den foreliggende oppfinnelse er å løse enhver syklisk eller ikke-syklisk avhengighet mellom forskjellige sikkerhetsdomener og som kan ha konsekvenser for de effektive brukerrettigheter til dokumentene.

En ytterligere hensikt med den foreliggende oppfinnelse er å minimere antall katalogsøk.

Nok en ytterligere hensikt med den foreliggende oppfinnelse er å skaffe en søkemotor som er i stand til å støtte og implementere fremgangsmåten i henhold til den foreliggende oppfinnelse.

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 gjenfinne aksesstillatelser fra brukerkataloger i multiple domener, idet et første domene av de multiple domener er avhengig av et annet domene blant disse hvis hovedbrukere i det første domene er dannet av brukere, grupper av brukere eller grupper som inneholder en eller flere nøstede og unøstede undergrupper, kan være hovedbrukere i det annet domene, å utlede domeneavhengigheter, å utlede en aksessekvens fra domeneavhengighetene, å aksessere brukerkatalogene med den utledede aksessekvens, å beregne filteret på grunnlag av aksesstillatelser for brukeren som stiller søkespørsmålet, å evaluere filteret i søkemotoren, å filtrere dokumentene returnert i resultatmengden, og å returnere dokumentene som har en aksesskontrolliste som tilsvarer filteret.

De ovennevnte hensikter så vel som ytterligere trekk og fordeler realiseres også med en søkemotor i henhold til den foreliggende oppfinnelse som er kjennetegnet ved at den omfatter en modul for å revidere søkespørsmålet slik at det gjenspeiler en nåværende brukers aksessadganger i kildedokumentmagasiner.

Ytterligere trekk og fordeler ved den foreliggende oppfinnelse vil fremgå av de vedføyde uselvstendige krav.

Den foreliggende oppfinnelse vil forstås bedre av den etterfølgende drøftelse av den generelle begreper og trekk så vel som fra drøftelser av eksempelutførelser derav ved at de refereres til konkrete applikasjoner og leses i samband med den vedføyde tegning, på hvilken

fig. 1 viser et eksempel på ikke-sykliske domeneavhengigheter,

fig. 2 et eksempel på sykliske domeneavhengigheter,

fig. 3 et eksempel på en nabomatrise for sykliske domeneavhengigheter,

fig. 4 et eksempel på en nabomatrise for et enkeltdomene,

fig. 5 et eksempel på en transitiv lukking av en nabomatrise for et enkeltdomene,

fig. 6 to eksempler på aktive katalogdomener og en lokal filtjenerdomene med brukere og grupper,

fig. 7 tre eksempler på aktive katalogdomener med brukere og grupper, og

fig. 8 skjematisk arkitekturen til en søkemotor i henhold til den foreliggende oppfinnelse.

Den generelle bakgrunn for den foreliggende oppfinnelse skal nå kort drøftes.

Fremgangsmåten i henhold til den foreliggende oppfinnelse kan anses om et ekstra verktøy eller raffinement som angår informasjonsaksess, søking og gjenfinning av informasjon over datakommunikasjonssystemer generelt, dvs. både ekstranett og intranett, hvor det foreligger en eller annen form for aksesskontroll påtvunget dokumentkildemagasinet. I denne kapasitet angår den søkemotorer hvor aksesskontroll i multiple domener gjennomføres før søkespørsmålsevalueringen ved å generere et såkalt forfilter. Dette filter evalueres som en del av søkespørsmålet ved å benytte aksesskontrollinformasjon som er blitt indeksert sammen med dokumentet. Følgelig må brukerens gruppemedlemskap i alle domener bestemmes, idet det tas hensyn til at den samme bruker eller gruppe kan forekomme i flere domener, direkte eller ved implisitt. Rett frem gjennomløping av medlemskapsgrafen vil kreve en rekke gjentatte katalogoppslag i flere domener.

Den foreliggende oppfinnelse angår både beskyttelsen av dokumenter og dokumentsammendragene og også oppdagelsen av alle relevante dokumenter i alle dokumentkildesystemer. Fremfor å benytte etterfiltreringsteknikker eller endre adgangskontrollmekanismene for eksisterende dokumentkildesystemer, viser den foreliggende oppfinnelse en fremgangsmåte som danner et søkefilter for den foreliggende bruker og som viser overensstemmelser bare hvis brukeren har aksess til de angjeldende dokumentene i kildesystemet. Følgelig vil resultatmengden fra søkespørsmålet være begrenset til dokumenter ved å iverksette midler og tiltak for å omskrive søkespørsmålet med et ytterligere filter.

Med andre ord er fremgangsmåten i henhold til den foreliggende oppfinnelse basert på å beregne et sikkerhetsfilter for hver bruker, basert på innhold i alle sikkerhetsdomenekataloger, og en beskrivelse av deres innbyrdes avhengigheter og avbildninger. Det beregnede sikkerhetsfilter svarer til en rad i en transitivt beregnet nabomatrise, fortrinnsvis i henhold til Warshalls algoritme som er kjent for fagfolk på området som en av de beste koder for å finne den transitive lukning av en graf og som starter fra nabomatrisen til grafen. Nabomatrisen til en rettet graf med n vertekser er n × n matrisen hvor hvert ikke-diagonal element a_{ij} er antallet kanter fra verteks *i* til verteks *j* og den diagonale element a_{ii} er antallet sløyfer i verteks *i*. Denne matrisen definerer i bunn og grunn grafen. Videre bør det bemerkes at den boolske nabomatrise er en nabomatrise hvor alle tall større enn 1 forandres til 1 og angir ikke avstanden, men i stedet oppnåelighet, dvs. begrepet for å være i stand til å komme fra en verteks til en eller annen annen verteks. Da bare én rad i Warshalls matrise er interessant på et gitt tidspunkt, kan forskjellige modifikasjoner av algoritmen benyttes. – For en mer omfattende drøftelse av nabomatriser og deres transitive lukning ved hjelp av Warshalls algoritme, kan det henvises til avsnitt 37.3.2 i J.K. Truss, Discrete Mathematics for Computer Scientists, Addison Wesley, New York, 1991.

Fremgangsmåten i henhold til den foreliggende oppfinnelse benytter en partiell ordning av domenene og en første breddegjennomløpning av disse for å garantere fullstendighet og minimal belastning på sikkerhetskatalogene samtidig som den frembringer resultatene fra Warshall-algoritmen. Som kjent for fagfolk vil en breddegjennomløping som utføres først, også kalt en første søk over bredden, være en grafsøkalgoritme som begynner ved rotnoden og utforsker alle nabonoder. For hver av de nærmeste noder utforsker den deretter disses uutforskede nabonoder og så videre, inntil den finner målet. Dette er forskjellig fra en dybde-først søk som starter ved roten og søker så langt som mulig langs hver gren før den går tilbake.

Dannelsen av et søkefilter i henhold til den foreliggende oppfinnelse skal nå forklares mer detaljert og med henvisning til tegningen. Fig. 1 viser et eksempel på ikke-sykliske domeneavhengigheter med skårer for optimal ordning, og fig. 2 et eksempel på sykliske domeneavhengigheter, tilsvarende skåret for optimal ordning. Først er det nødvendig med en beskrivelse av alle sikkerhetsdomener D og deres avhengigheter M som er en liste av relasjoner D×D.

For hvert domene $d \in D$ må det da foreligge en definert brukerovervåker UM_d som for hver bruker $u \in U_d$ kjenner foreldergruppene $g \in G_d$ som brukeren er medlem av. Unionen $P_d = U_d \cup G_d$ kalles prinsipalene i ett sikkerhetsdomene og inneholder alle brukere og grupper i sikkerhetsdomenet. Her kan en gruppe være gruppe av brukere eller gruppe med undergrupper som er inneholdt nøstet eller unøstet i gruppen. P er definert som en union av alle P_d og er mengden av alle brukere og grupper i alle sikkerhetsdomener.

En forelderfunksjon er gitt som

$$Parent_d: P_d \rightarrow P_d^*$$

For hver domeneavhengighet $m \in M$ mellom domener $i \in D$ og $j \in D$, kreves det at det er en kryssdomeneoppløser som kjenner funksjonen

$$Alias_{i,j} : P_i \rightarrow P_j^*$$

Basert på det ovennevnte kan det settes opp en nabomatrise A slik at en del av matrisen kommer fra brukerovervåkere (forelderfunksjonen) og resten fra kryssdomeneoppløsere (aliasfunksjonen). Som ovenfor nevnt, er sykliske domeneuavhengigheter med skåre for optimert ordning vist på fig. 2. Fig. 3 viser et eksempel på hvordan avhengigheter for domener i figur 2 avbildes til nabomatrisen. På fig. 3 representerer hver rad og kolonne flere rader og kolonner i den virkelige nabomatrise, én for hver prinsipal i domenet som benytter Warshalls algoritme.

Nå må den transitive lukning TC av A bestemmes. Den transitive lukning av en rettet graf er oppnålighetområdet til grafen. For en rettet graf med n vertekser, vil det foreligge en n × n-matrise som beregnes som

$$TC(A) = I + A + A^2 + A^3 + \ldots A^n$$

hvor n kan være et hvert tall opp til $|P|$.

Hver gang en bruker u gjennomfører et søk, er det bare nødvendig med en rad av TC(A), nemlig raden som svarer til denne bruker. Det er derfor unødvendig å beregne hele TC(A), men bare de deler derav som er relevante for resultatet av rad u.

Før en rad av TC(A) beregnes, må den orden hvormed domenene besøkes bestemmes ved å utføre følgende trinn.

a) Beregne en skåre for hvert domene basert på hvor mange domener som kan nås fra dette i avhengighetsgrafen. Igjen kan det henvises til eksemplene på fig. 1 og fig. 2.

b) Sorter domenene etter fallende skåre.

For å beregne en enkelt rad av TC(A) svarende til bruker u, må deretter følgende trinn utføres.

a) Start med en initialt tom mengde av prinsipaler R.

b) For hvert domene d, dann en initialt tom mengde av prinsipaler L_d.

c) Adder brukeren u til mengden av prinsipaler L_d for domener d hvor u er definert.

Nå skal følgende undertrinn gjentas inntil L_d er tom for alle domener d.

a) Velg det første domenet d (basert på den forhåndsberegnede skåre) med en ikke-tom L_d.

b) Adder prinsipalen i L_d til R.

c) La M være unionen av $Parent_d(p)$ for alle prinsipaler p i L_d.

d) Slett L_d.

e) Adder prinsipalene i M til R.

f) For alle etterfølgere s av d i avhengighetsgrafen og alle prinsipaler m i M, beregn $Alias_{d,s}(m)$ og adder til L_s.

R inneholder nå alle grupper som brukeren u er medlem av. Den ønskede rad i TC(A) inneholder en innførsel 1 for alle prinsipaler i R og 0 for alle andre.

Hvis det ikke foreligger noen sykler i avhengighetsgrafen, besøkes hvert domene bare en gang. Hvis det forekommer sykler, vil domener med sykliske avhengigheter få samme skåre og kan besøkes om igjen i trinn a) straks ovenfor inntil det ikke oppdages noen flere foreldre i noen av disse domener.

En enkel nabomatrise A for et enkelt domene med en bruker "john" er vist på fig. 4. "john" er et medlem av gruppen "hr", som igjen er et medlem av "admin". Den transitive lukning av dette vil være som vist på fig. 5. Det skal

bemerkes at raden med "john" viser at han direkte eller indirekte er medlem av både "hr" og "admin".

Gitt at denne ene rad av TC(A) som svarer til den foreliggende bruker kan nå et søkefilter konstrueres ved å addere en disjunksjon av brukerens gruppemedlemskap, for eksempel slik:

SØKEEKSEMPEL: test or "foo bar"

BRUKER: john

FORELDER TIL BRUKER: hr, admin

RESULTERENDE SØK: (test or "foo bar") and (docacl:john or docacl:hr or docacl:admin)

Hvis ACL-feltet i dokumentet (kalt docacl) også kan inneholde bannlyste brukere hvor en "9" foran impliserer at vedkommende er bannlyst, kunne det resulterende søkespørsmål være som dette:

RESULTERENDE SØK: (test or "foo bar") and (docacl:john or docacl:hr or docacl:admin) andnot docacl:9john andnot docacl:9hr andnot docacl:9admin

En rekke eksemplifiserende utførelser av den foreliggende oppfinnelse skal nå gis uttrykt ved spesifikke applikasjoner derav.

Eksempel 1

I en anbringelse typisk for et stort foretak er det en mengde fallgruver med Active Directory™ og tillatelser. Det er f.eks. mulig å danne lokale grupper som inneholder universelle brukere som medlemmer på en filtjener. Disse lokale grupper kan så benyttes til å utstede tillatelse på filer på denne filtjeneren. Når imidlertid gruppemedlemskapene til en bruker bestemmes mot den globale katalog eller domenekontoller for brukeren, kan dennes gruppemedlemskap på filtjeneren ikke gjenfinnes. Så det er også nødvendig å spørre filtjeneren om gruppemedlemskaper i denne og kombinere disse resultatene. En lignende situasjon oppstår med domenelokale grupper.

Den nye tilnærming løser dette problemet ved ganske enkelt å beskrive alle domener (og å beskrive en filtjener som et domene), deres lenker og hvilke brukerovervåkere og kryssdomeneresolvere som henholdsvis kjenner gruppemedlemskapene (forelderfunksjonen) og avbildningen mellom domenene (aliasfunksjonen).

Fig. 6 viser et forenklet eksempel av dette scenario med tre domener. To av domenene er aktive katalogdomener (domene 1 og domene2), mens det tredje domene er en filtjener med lokale brukere og grupper. Bruker u_5 i domenet 1 har et alias i domene 2, som er et medlem av to grupper (g_{11} og g_{12}) i domenet 2. Gruppe g_{11} i domene 2 har et alias i domenet 3 som er et medlem av en lokal gruppe (g_{21}) på filtjeneren. For å bestemme brukeren fullstendig, må alle tre domener besøkes.

Eksempel 2

En annen utførelse av den foreliggende oppfinnelse foreliggere innenfor intranettsøking med innbyrdes sykliske domener. I et slikt scenario kan det være nødvendig å besøke hvert domene flere ganger for å bestemme en bruker fullstendig. Fig. 7 illustrerer dette eksempel. På figuren er det tre Active Directory™-domener, et forelderdomene og to underdomener. Den sykliske avhengighet eksemplifiseres av aliasene mellom domene 2 og domene 3. For å bestemme at brukeren u_1 er et medlem av g_{13} (så vel som g_1, g_3, g_{11}, g_{12} og g_{21}), må domene 2 besøkes to ganger da det foreligger en syklisk avhengighet.

Et generelt system for aksessering, søking og gjenfinning av informasjon hvor fremgangsmåten i henhold til den foreliggende oppfinnelse kan anvendes, kan det fordelaktig være utført på en søkemotor i henhold til den foreliggende oppfinnelse.

I det følgende skal en søkemotor innrettet for å støtte og implementere fremgangsmåten i henhold til den foreliggende oppfinnelse drøftes i noe detalj. For å støtte og implementere fremgangsmåten på den foreliggende oppfinnelse er ytterligere komponenter og moduler anordnet og skal beskrives med henvisning til fig. 8.

Søkemotoren 100 i henhold til den foreliggende oppfinnelse skal som kjent i teknikken omfatte forskjellige undersystemer 101-107. Søkemotoren kan aksessere dokument- eller innholdsmagasiner som befinner seg i et innholdsdomene eller -rom, hvorfra innholdet enten aktivt skyves inn i søkemotoren eller via en datakobling trekkes inn i søkemotoren. Typiske magasiner innbefatter databaser, kilder som står til rådighet via ETL (Extract-Transform-Load) -verktøy slik som Informatica, ethvert XML-formatert magasin, filer fra filtjenere, filer fra vevtjenere, dokumenthåndteringssystemer, innholdshåndteringssystemer, e-postsystemer,

kommunikasjonssystemer, samarbeidssystemer og rike media, så som audio, bilder og video. Magasinene kan tilhøre forskjellige sikkerhetsdomener. Hvert dokument inneholder en ACL (Access Control List) som definerer grupper og brukere som har aksess til dokumentet. De gjenfunne dokumenter leveres til søkemotoren 100 via en innholds-API (Application Programming Interface) 102. Deretter blir dokumentene analysert i et innholdsanalysetrinn 103, også betegnet et undersystem for forhåndsprosessering av innhold for å klargjøre innholdet for forbedrede søke- og oppdagelsesoperasjoner. Utgangen fra innholdsanalysen benyttes til å mate kjernesøkemotoren 101.

Kjernesøkemotoren 101 kan typisk være anbrakt over en tjenerfarm på en desentralisert måte for å tillate prosessering av store dokumentmengder og høye spørsmålsbelastninger. Kjernesøkemotoren 101 kan motta brukeranmodninger og frembringe lister av overensstemmende dokumenter. I tillegg kan kjernesøkemotoren 103 frembringe ytterligere metadata for resultatmengden, f.eks. sammendragsinformasjon for dokumentattributter.

Kjernesøkemotoren 101 omfatter i seg selv ytterligere undersystemer, nemlig et indekseringsundersystem 101a for nedsamling ("crawling") og indeksering av innholdsdokumenter og et søkeundersystem 101b for å utføre den egentlige søking og gjenfinning. Alternativt kan utgangsdataene fra innholdsanalysetrinnet 103 mates inn i en valgfri varslingsmotor 104. Varslingsmotoren 104 vil ha lagret en mengde søkespørsmål og kan bestemme hvilke søkespørsmål som ville ha akseptert den gitte dokumentinnmating. En søkemotor kan aksesseres fra mange forskjellige klienter eller applikasjoner som typisk kan være mobile og datamaskinbaserte klientapplikasjoner. Andre klienter innbefatter PDAer og spillinnretninger. Disse klientene, som befinner seg i et klientrom eller -domene, vil levere anmodninger til en søkemotor-API 107, nemlig et søkespørsmål eller klient-API. Søkemotoren 100 vil typisk besitte et ytterligere undersystem i form av et søkespørsmålsanalysetrinn 105 for å analysere og forfine søkespørsmålet med tanke på å konstruere et avledet søkespørsmål, hvilket er det som i realiteten eksekveres ved hjelp av søkemotoren 101. Formålet med denne forfining kan være å ekstrahere mer meningsfylt informasjon, eller slik det er tilfellet ved denne oppfinnelsen, å forsyne søkespørsmålet med systemdefinert sikkerhetsrutiner. Således kan dette undersystem innbefatte en sikkerhetsomformer 108 som er ansvarlig for å generere et sikkerhetsfilter for brukeren som utsteder søkespørsmålet. Endelig kan utgangsdataene fra

kjernesøkemotoren 101 typisk ytterligere analyseres i et annet undersystem, nemlig et resultatanalysetrinn 106 for å frembringe informasjon eller visualiseringer som benyttes av klientene. Dette undersystem kan også innbefatte en ettersikkerhetsfiltreringsprosess som er ansvarlig for å verifisere at brukeren har aksess til dokumentene i søkeresultatet ved å kommunisere med dokumentmagasinet. – Begge trinn 105 og 106 er forbundet mellom kjernesøkemotoren 101 og klient-API 107, og i tilfelle varslingsmotoren 104 foreligger, er den forbundet i parallell med kjernesøkemotoren 101 mellom innholdsanalysetrinnet 103 og søkespørsmåls og resultatanalysetrinnene 105;106.

For å støtte og implementere den foreliggende oppfinnelse må søkemotoren 100 som kjent i teknikken være utstyrt med en modul 108 som svarer til sikkerhetsomformeren. Modulen 108 er anordnet i søkespørsmålsanalysetrinnet. Alternativt kan modulen 108 være anbrakt i kjernesøkemotoren 101 og utføre den samme funksjon.

Den foreliggende oppfinnelse viser hvordan aksesstillatelser for en bruker som utsteder et søkespørsmål effektivt kan finnes i et miljø med flere avhengige sikkerhetsdomener og skaffer en løsning på de utfordringer slike domener representerer, samtidig som det benyttes eksisterende infrastrukturer for sikkerhetsdomener uten å måtte ty til etterfiltrering. Ved å evaluere avhengighet mellom sikkerhetsdomener og finne den optimal orden av domener, blir forsinkelsen ved generering av sikkerhetsfilter minimert og den oppfattede kvalitet av søkemotoren økes. Ved å prosessere interdomeneavhengigheter, vil fremgangsmåten i henhold til den foreliggende oppfinnelse dessuten unngå å utføre potensielt kostbar etterfiltrering av dokumenter, og følgelig kan søkespørsmålsytelsen i en desentralisert søkemotor økes. Avhengighetene mellom domener blir brukt til ytterligere å avgrense søket og unngå oppslag i domener som ikke kan bidra, og spesielt gjelder det gjentatte besøk til det samme domene.

Den foreliggende oppfinnelse representerer således en betydelig forbedring i forhold til de vanlige benyttede metoder for dokumentautorisasjon ved aksess, søking og gjenfinning av informasjon, slik det vil fremgå av det ovenstående.

PATENTKRAV

1. Fremgangsmåte for å begrense aksess til søkeresultater i form av dokumenter hentet fra et dokumentmagasin, hvor fremgangsmåten angår et system for aksessering eller søking av informasjon, hvor en bruker av systemet for søking eller aksessering benytter et søkespørsmål på dokumentmagasinet for å innhente en resultatmengde i form av dokumenter fra dette, hvor aksessen er begrenset til de dokumenter i resultatmengden eller alle gjenfunne dokumenter som har en aksesskontrolliste som tilsvarer et filter utført som et søkespørsmål, hvor systemet for aksess eller søking av informasjon er implementert på en søkemotor, og hvor fremgangsmåten er
k a r a k t e r i s e r t v e d
å gjenfinne aksesstillatelser fra brukerkataloger i multiple domener, idet et første domene av de multiple domener er avhengig av et annet domene blant disse hvis hovedbrukere i det første domene er dannet av brukere, grupper av brukere eller grupper som inneholder en eller flere nøstede og unøstede undergrupper, kan være hovedbrukere i det annet domene,
å utlede domeneavhengigheter,
å utlede en aksessekvens fra domeneavhengighetene,
å aksessere brukerkatalogene med den utledede aksessekvens,
å beregne filteret på grunnlag av aksesstillatelser for brukeren som stiller søkespørsmålet,
å evaluere filteret i søkemotoren,
å filtrere dokumentene returnert i resultatmengden, og
å returnere dokumentene som har en aksesskontrolliste som tilsvarer filteret.

2. Fremgangsmåte i henhold til krav 1,
k a r a k t e r i s e r t v e d å beskrive domeneavhengigheter eksplisitt, og å gjøre dem tilgjengelig som inndata for å utlede aksessekvensen.

3. Fremgangsmåte i henhold til krav 2, hvor domeneavhengighetene danner en partiell orden,
k a r a k t e r i s e r t v e d å besøke domener i en topologisk sortert orden, slik at hvert domene blir besøkt høyst en gang.

4. Fremgangsmåte i henhold til krav 2, hvor domeneavhengighetene har sykler,
k a r a k t e r i s e r t v e d å oppdage sykliske avhengigheter ved å identifisere minimale sykler, og

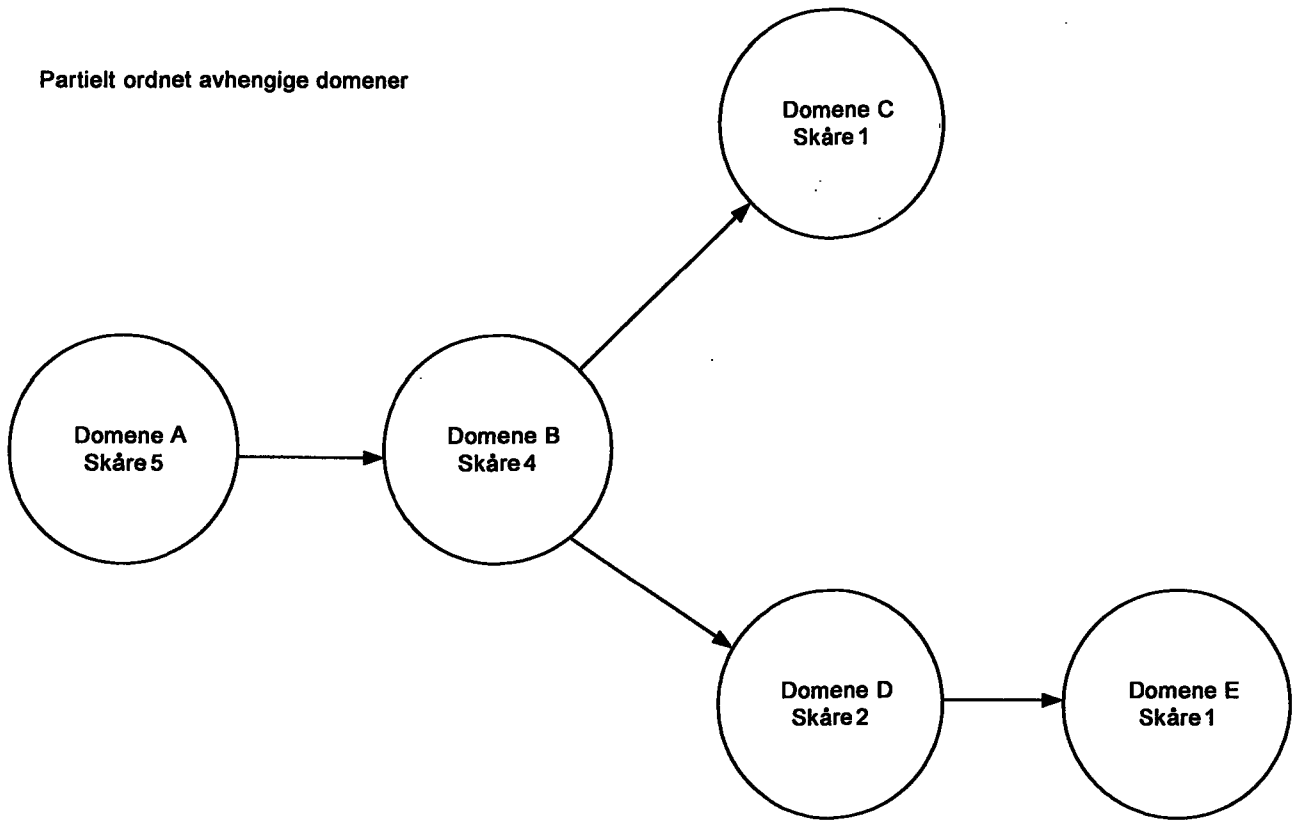
å iterere over de angjeldende domener inntil ingen ytterligere grupper tilføyes til en mengde av aksesstillatelser.

5. Søkemotor (100) som er i stand til å støtte og implementere fremgangsmåten i henhold til et av de foregående krav i systemer for aksess eller søking av informasjon, hvor søkemotoren (100) benyttes til å aksessere, søke, gjenfinne og analysere informasjon fra dokument- eller innholdsmagasiner som er tilgjengelige via datakommunikasjonsnettverk, innbefattet ekstranett og intranett, og å fremlegge søke- og analyseresultater for sluttbrukere, hvor søkemotoren omfatter minst en kjernesøkemotor (101) et applikasjonsprogrammert grensesnitt for innhold (102) (innholds-API) forbundet med den minst ene kjernesøkemotor (101) via et innholdsanalysetrinn (103), og et applikasjonsprogrammert grensesnitt (107) for søkespørsmål forbundet med den minst ene kjernesøkemotor (101) via respektive søkespørsmålsanalyse- og resultatanalysetrinn (105;106), og hvor søkemotoren er
k a r a k t e r i s e r t v e d at den omfatter en modul (108) for å revidere et søkespørsmål til å gjenspeile en nåværende brukers aksesstillatelser i kildedokumentmagasiner.

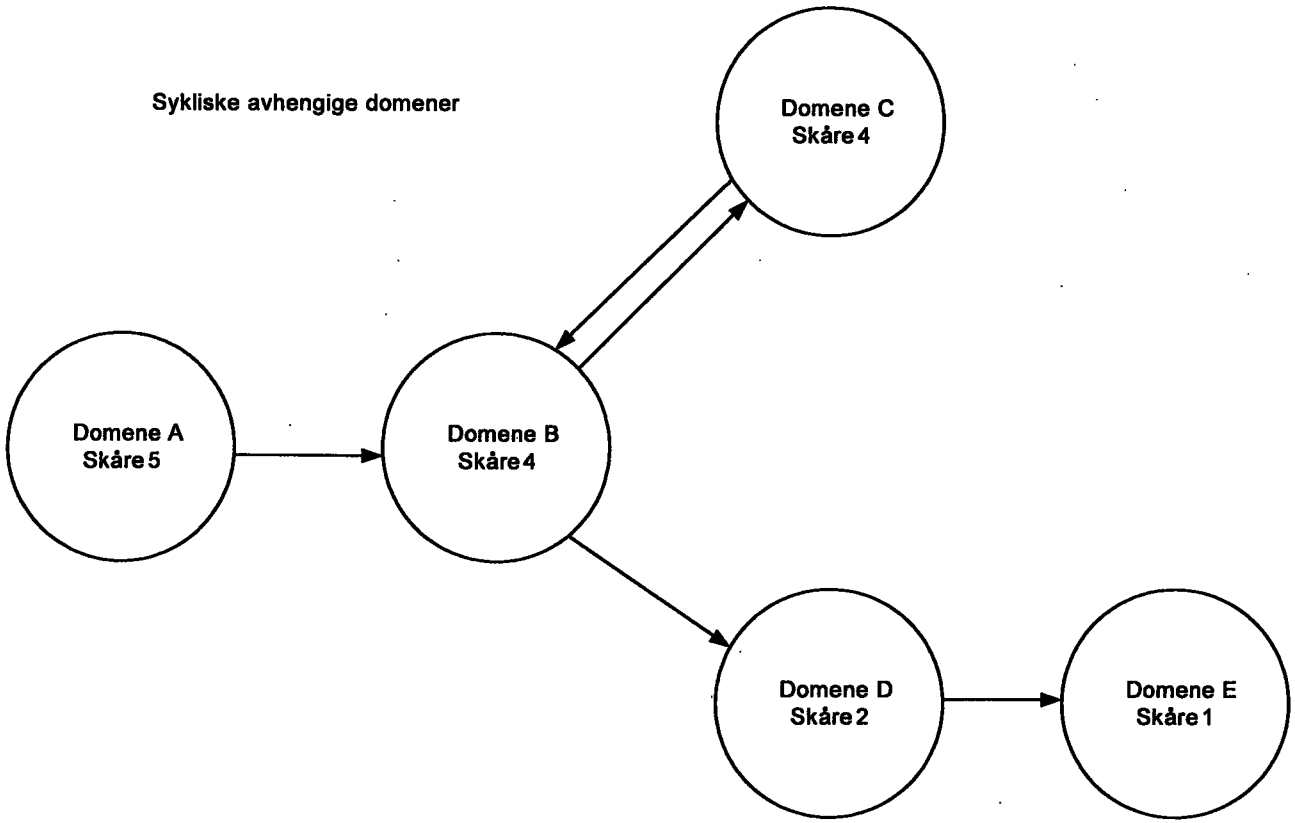
6. Søkemotor (100) i henhold til krav 5,
k a r a k t e r i s e r t v e d at modulen (108) er anordnet i søkespørsmålsanalysetrinnet (105).

7. Søkemotor (100) i henhold til krav 5,
k a r a k t e r i s e r t v e d at modulen (108) er anordnet i den minst ene kjernesøkemotor (101).

8. Søkemotor (100) i henhold til krav 5,
k a r a k t e r i s e r t v e d at modulen (108) er innrettet til å revidere søkespørsmålet som et sikkerhetsfilter for en nåværende bruker.


Partielt ordnet avhengige domener
Domene C
Skåre 1
Domene A
Skåre 5
Domene B
Skåre 4
Domene D
Skåre 2
Domene E
Skåre 1

Figur 1


Sykliske avhengige domener
Domene C
Skåre 4
Domene A
Skåre 5
Domene B
Skåre 4
Domene D
Skåre 2
Domene E
Skåre 1

Figur 2

	Prinsipaler i Domene A	Prinsipaler i Domene B	Prinsipaler i Domene C	Prinsipaler i Domene D	Prinsipaler i Domene E
Prinsipaler i Domene A	$Forelder_A$	$Alias_{A,B}$	0	0	0
Prinsipaler i Domene B	0	$Forelder_B$	$Alias_{B,C}$	$Alias_{B,D}$	0
Prinsipaler i Domene C	0	$Alias_{C,B}$	$Forelder_C$	0	0
Prinsipaler i Domene D	0	0	0	$Forelder_D$	$Alias_{D,E}$
Prinsipaler i Domene E	0	0	0	0	$Forelder_E$

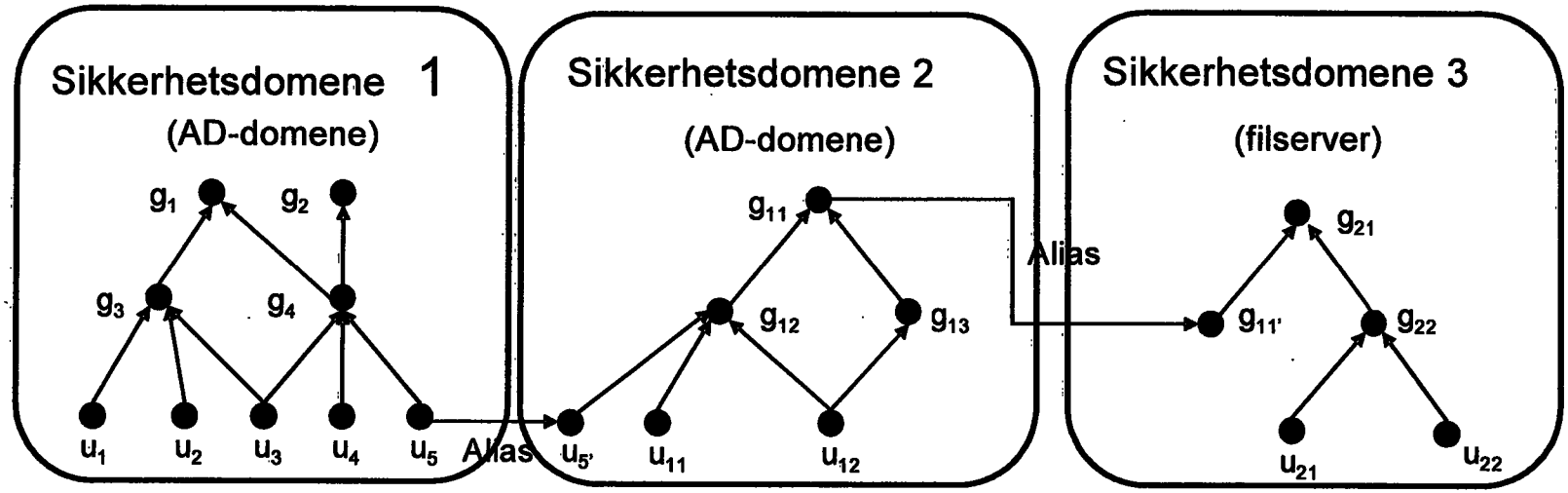
Figur 3

	john	hr	admin
john	1	1	0
hr	0	1	1
admin	0	0	1

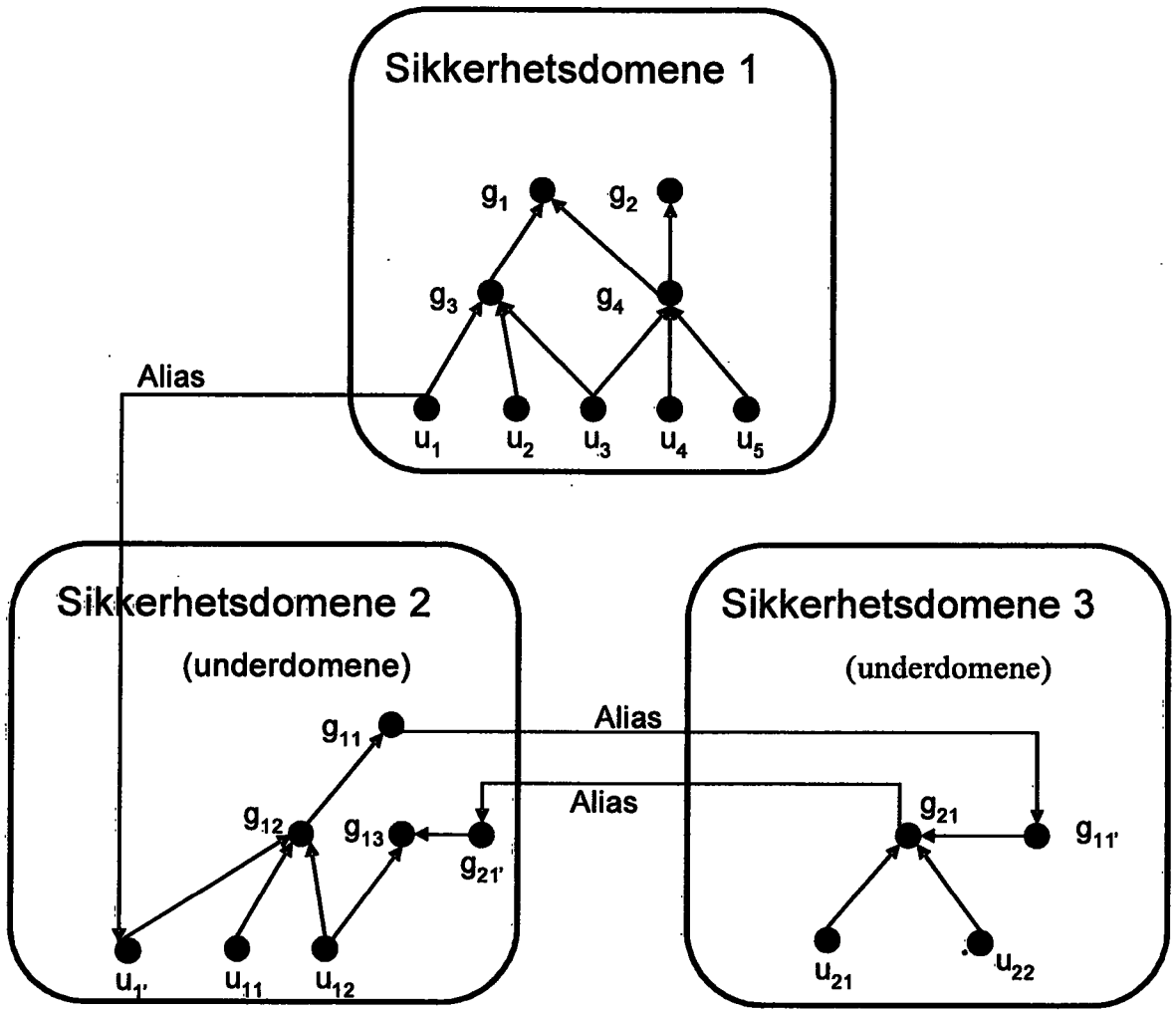
Figur 4

	john	hr	admin
john	1	1	1
hr	0	1	1
admin	0	0	1

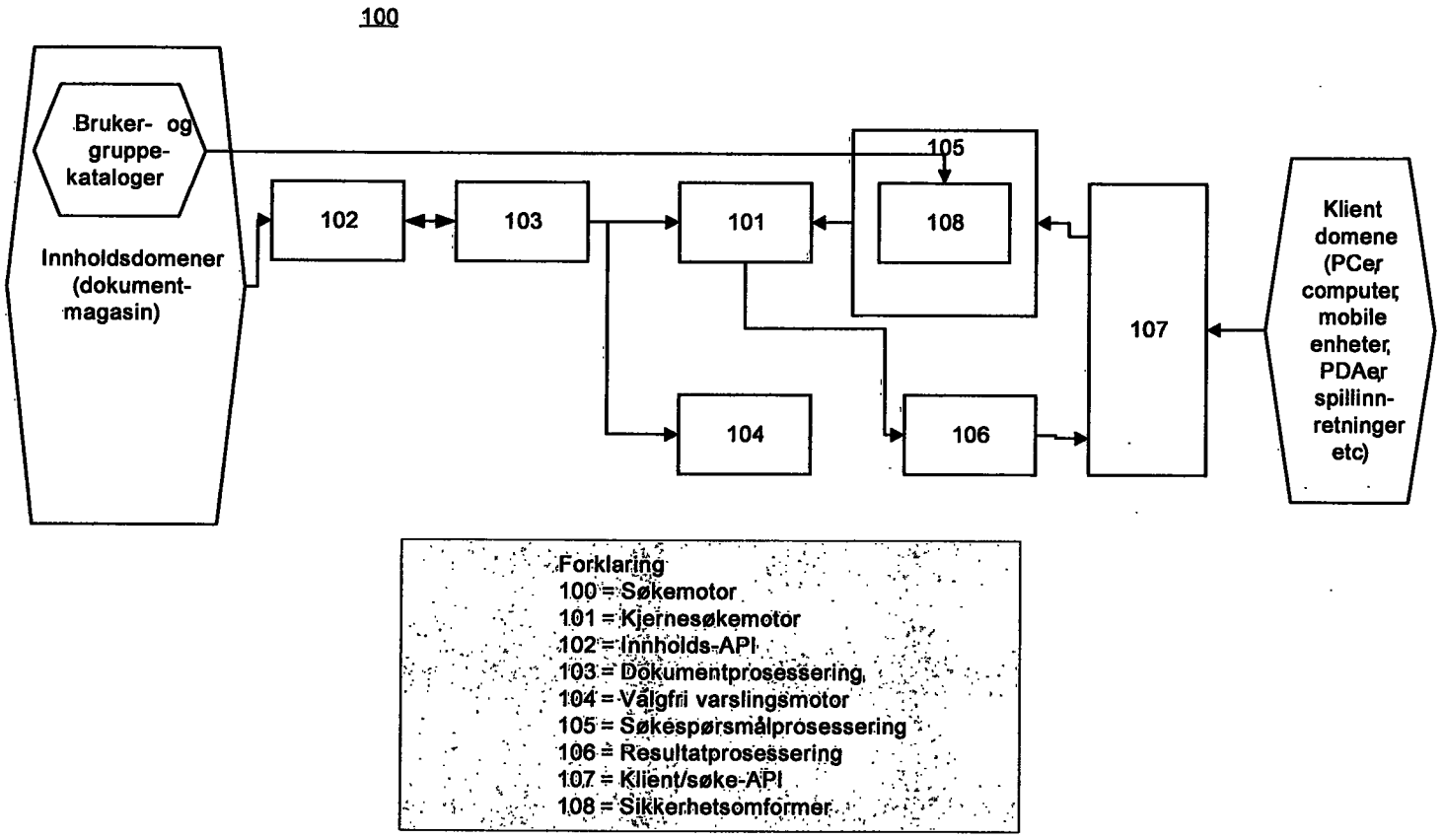
Figur 5


Sikkerhetsdomene 1
(AD-domene)
g1
g2
g3
g4
u1
u2
u3
u4
u5
Alias
Sikkerhetsdomene 2
(AD-domene)
u5'
g11
g12
g13
u11
u12
Alias
Sikkerhetsdomene 3
(filserver)
g21
g11'
g22
u21
u22

Figur 6


Sikkerhetsdomene 1
g1
g2
g3
g4
Alias
u1
u2
u3
u4
u5
Sikkerhetsdomene 2
(underdomene)
Sikkerhetsdomene 3
(underdomene)
g11
Alias
Alias
g12
g13
g21'
g21
g11'
u1'
u11
u12
u21
u22

Figur 7


100
Bruker- og gruppe-kataloger
Innholdsdomener (dokument-magasin)
102
103
101
105
108
107
104
106
Klient domene (PCer, computer, mobile enheter, PDAer, spillinn-retninger etc)
Forklaring
100 = Søkemotor
101 = Kjernesøkemotor
102 = Innholds-API
103 = Dokumentprosessering
104 = Valgfri varslingsmotor
105 = Søkespørsmålprosessering
106 = Resultatprosessering
107 = Klient/søke-API
108 = Sikkerhetsomformer

Figur 8