

Gépi indexelés

Karen SPARCK JONES

A Journal of Documentation rendszeresen megjelenő szemléi sorában 1974-ben jelent meg a téma egyik legjobb kutatója tollából a gépi indexelés helyzetét, utolsó öt évi fejlődését összefoglaló tanulmány, amelyet az alábbiakban ismertetünk. (Automatic indexing. Journal of Documentation, 1974. 4.no. 393-432.p. bibliogr. 105 tétel) A szemlét SÁRDY Péter tömörítette.

A TÉMA ÉS KÖRNYEZETE

A gépi indexelés kezdeteként Luhn 1958-as gépi referátumai tekinthetők. A jelenlegi helyzetet illetően a vélemények megoszlanak. A szakirodalom alapján úgy tűnik, hogy a cimből kiinduló egyszerű gépi eljárásokat gyakorlati okokból elfogadták, de a bonyolultabb módszerek még nem jutottak túl a kutatási fázison.

Az indexelés az információvisszakereső rendszerek nyelvi összetevőivel foglalkozik: dokumentumok és kérdések szövegelemzése, leírása, esetleges módosítása, továbbá az indexnyelv kialakítása. Elvileg mind a négy tevékenység automatizálható, a valóságban azonban rendszerint csak részben gépesített rendszerek találhatók.

A probléma pontosabb leírásához el kell választani annak szemantikai és szintaktikai oldalát.

A visszakereső rendszer szemantikájához tartozik

1. a dokumentum/kérdés feldolgozása felől

- a) elemzés, szavak és egyszerű fogalmak meghatározása;
- b) leírás, indexkifejezések vagy deszkriptorok megadása;
- c) visszakeresés, a kifejezések közti paradigmatis vagy osztálykapcsolatok feltárása;

2. az indexnyelv felől

- a) a szótár megalkotása, kifejezések, deszkriptorok megválasztása;
- b) a szótár szerkesztése, paradigmatis kapcsolatok létrehozása.

A szintakszishoz tartozik

3. a dokumentum szempontjából

- a) elemzés, szintaktikai szerepek és kapcsolatok azonosítása;
- b) leírás, kapcsolat-információk biztosítása;
- c) visszakeresés, kifejezések közti szintagma-relációk feltárása;

4. a nyelv szempontjából

- a) a nyelv megalkotása, a kifejezések vagy deszkriptorok összekötése szintagma- vagy logikai kapcsolók megválasztása;
- b) a nyelv szervezése, szükséges modulációs szabályok megalkotása.

A fenti részek közül csupán a 4. pontban felsoroltakra nincs gépi megoldási példa.

A gépi indexelés "környezetéből" elsősorban az alábbi témakörök veendőek figyelembe: visszakereső rendszerek (legjelentősebb fejlődés technikai téren); visszakeresési kísérletek (a nagyszámú kísérlet még mindig nem vezetett megfelelő elmélet kialakításához); számítástechnika; nyelvtudomány (a közelmúlt legfontosabb eredménye a szöveg felszine és mélystrukturája közti elhatárolás, de még hiányzik a megfelelő transzformációs nyelvtan); faktografikus visszakeresés (egyelőre kísérleti stádiumban).

A gépi indexelés értékelése kétféleképp történhet: külsőleg (manuális eljárásokkal összevetve) vagy belsőleg (más gépi eljárásokkal összevetve). Megkülönböztethetünk makro (a rendszer egészére vonatkozó) és mikro (egy konkrét elemre vonatkozó) értékelést. Sajnos a jelenleg használt értékelési eszközök nem mindig megfelelőek, de két mérték, a visszahívás és pontosság oly széles körben alkalmazott, hogy elméleti korlátaik ellenére sem mellőzhetők. Az értékelés kiindulópontja mindenképpen a visszakeresési teljesítmény kell legyen.

A téma tárgyalása során mindig figyelemmel kell lenni néhány releváns tényezőre, mint pl. gyűjteményi jellemzők (méret, dokumentum-típus, nyelv, tárgy), működési módok (többszörös szolgáltatások, költség-minimalizálás), input-szöveg megválasztása, a relevancia értelmezése, az összehasonlítás típusa stb.

SZINTAKSZIS

A dokumentum leírásában a szintaktikai információ vagy implicit, vagy explicit. Az első esetben összevetéskor csak a jelenlévő tárgyszavak kerülnek felhasználásra, a másodikban kapcsolataik is. Az egyszerű koordinált indexelés olyannak tekintendő, mint amely nem hordoz szintaktikai információkat, s a kérdés az, hogy mennyi ilyen információra van szükség, s ezek biztosíthatók-e gépi uton.

A kétféle szintaktikai elemzésnek megfelelően az input szövege elemzésének vagy szó-láncok kialakítása a célja, vagy a szöveg kapcsolat-szerkezetének kifejtése. A valóságban természetesen e két szélső eset közt számos átmeneti fázis van, amelyek a szerkezeti elemzés fokában térnek el egymástól.

A részleges szintaktikai elemzés legalsóbb foka a különféle elemek kiválogatásához használt speciális jelek és jelsorok (pl. írásjelek, prepozíciók, kötőszók stb.) alkalmazása. Ez esetben a meghatározásra kerülő szó-láncok szintaktikai jellemzésére nem kerül sor, s a gépi eljárás egyes esetekben csak előkészíti a manuális indexelést.

Középső szintnek tekinthető az az eljárás, amelynek során a mondat részleges vagy teljes elemzése megtörténik, de a strukturális leírás foka alacsony. Az e típusban

eddig lefolytatott kísérleteknél különféle hiányosságok észlelhetők (a lehetőségek hiányos kihasználása, az alkalmazott eszközök ad hoc jellege, stb.).

Teljesnek tekinthető elemzésre eddig csak a Smart rendszer korai változatában volt példa. Az ott alkalmazott eljárás azonban igen költségesnek, s ehhez képest meglehetősen eredménytelennek bizonyult.

Történt néhány kísérlet a statisztikai szintaxis területén is, azonban a kiinduló feltevések eddig gyengének bizonyultak érdemleges eredmények eléréséhez.

Az elemzési eljárások néhány általános jellemzőjét érdemes megemlíteni: információ az input szövegből; a leírás során való alkalmazhatóság hiánya; hatékonysági elemzések bizonytalansága.

Az input szöveg szintaktikai elemzésével szemben a dokumentum leírásának szintaxisában az explicit megoldásra nincs kísérlet. Az első lépés során kiválasztott kifejezéseket, szóláncokat rendszerint nem-automatikus módszerekkel kezelik.

A szintaktikai funkciók kifejezésére egyetlen lényeges kísérlet történt: a Syntol 1. Ennek leírása szintagmákból áll (általános relációkkal összekapcsolt kifejezés-párok). Az eljárást manuális indexeléssel összevetve 97 % helyes eredményt találtak, ezzel szemben azonban az index-kifejezések 20 %-a izolált deskriptor, s a szintagmák 80 %-át az igen gyenge "asszociatív" kapcsolat tartja össze.

Az indexnyelv szintaktikáját illetően számos elméleti kutatás folyt és folyik, gyakorlatilag azonban nem került sor az automatizálás lehetőségeinek vizsgálatára, mégkevésbé kísérleti alkalmazására.

Hasonló megállapítás érvényes a keresési szintaxisra is, ami az automatikusan előállított összetett deskriptorok keresési felhasználását illeti. Explicit szintaktikai információ esetén ez a keresés során módosítható, de az ehhez szükséges rendszeres szabályokat láthatóan nem nagyon tanulmányozták még. A Syntol korai szakaszában próbálkoztak a szintagmán belüli kapcsolatok modulálásával, az újabb változatok azonban ezt nem tartalmazzák.

Természetes, hogy amennyiben a dokumentum leírásában szerepelnek szintaktikai információk, hasonlókat kell biztosítani a kérdések feldolgozásában is. Az automatizálásra itt is kevés példa akad, de a manuálisan kialakított Boole-kapcsolatok szintén szintaktikai jellegűeknek tekinthetők.

A szintaktikai információ visszakeresési értéke erősen vitatott. A világosabb kép kedvéért bizonyos megkülönböztetésekre van szükség. pl. a szintaktikai kritérium (szintaktikus információ alkalmazása szemantikailag fontos indexelési elemek kiválasztására) és a szemantikai információ közt. Az első esetben a kérdés, hogy az így kiválasztott elemek jobbak-e, mint a másként, pl. statisztikailag kiválogatottak. A kérdés eldöntésére sajnos még nincsenek megfelelő tapasztalatok. A második esetben a kérdés, az, hogy a szintaktikai információ hasznos-e, szükség van-e explicit információra, s hogy a probléma automatizálható-e. Manuális példák arra mutatnak, hogy általában egyik típusú információ sem részesíthető előnyben. Az automatizálásra ill. annak hatékonyságára vonatkozóan még számos kísérletnek kell megelőznie a megnyugtatónak ítéltető választ.

Az információvisszakeresésben az automatizált szemantika legnagyobb előrelépése eddig a tartalmi leírás statisztikai megközelítése volt. Ez lehet oka annak is, hogy a szemantikai elemzés és leírás más lehetséges technikái csaknem teljesen hiányoznak.

Az input szemantikájának célja a tartalmat hordozó szavak meghatározása. Az elemzés null-szintjét a teljes szöveg vagy referátum tárolása jelenti.

A szintaktikai módszereket nem alkalmazó szó-válogatás vagy szótáron, vagy statisztikai módszereken alapul. Az ilyen értelmű elemzések során elsősorban negatív szótárakat alkalmaznak. A statisztikai módszerek alkalmazásának előfeltétele a megfelelően hosszú szöveg. Az alapfeltételt Luhn fogalmazta meg, eszerint a szógyakoriság durván arányos a szó szemantikai fontosságával. A probléma rendszerint a gyakoriság meghatározása és a megfelelő küszöb megválasztása. Az előbbi a nehezebb, hiszen a szövegben előfordulnak szinonimák is, és a szó előfordulási gyakorisága önmagában néha semmit sem mond. Emiatt kísérleteznek különféle kiegészítő eszközökkel, elsősorban a relatív gyakoriság alkalmazásával. Az utóbbi időben e téma viszonylag kevésbé haladt előre, elsősorban a fellépő igen magas költségek miatt. Tulajdonképpen csak egy, a megszokottól eltérő kísérlet folyt le, Cagan vezetésével, amelyben a szavak páronkénti előfordulási gyakorisága adta az elemzés alapját.

A dokumentum-leírás szintetikájánál a legfontosabb probléma a szövegből származó kifejezések felhasználása. A két legjellemzőbb lehetőség: a derivatív indexelés, amely a nyert kifejezéseket változatlanul használja, ill. a kijelöléses indexelés, ahol a szövegből származó kifejezéseket szükség esetén szótárból történő kijelöléssel helyettesítik. A gépi indexelés alapvetően mindig derivatív, de a leírás során lehetséges megfelelő szótár alkalmazása.

A kulcsszó normalizálásának jelentősége általánosan elismert. A szóvégi ragok, képzők stb. automatikus levágásának lehetőségét a nyelvészek már vizsgálták. Optimális eredmények csak megfelelő szótárak használatával látszanak elérhetőnek, ez azonban rendkívül költséges volna. Bizonyos hiba-arány még így sem kerülhető el, ez azonban nem hat ki lényegesen a visszakeresés hatékonyságára.

Egy indexnyelvből kétféleképpen adhatók meg indexkifejezések a dokumentumhoz: különböző input-szavakat azok osztálya nevével helyettesítve ill. egyes szavakat szóhalmazokkal helyettesítve. Az első esetben a deskriptorok a visszahívást, a másodikban a pontosságot erősítik. Bizonyos indexelési rendszerek a két eljárásnak csak egyikét teszik alkalmazhatóvá, de vannak gépi eredetű nyelvek, amelyekkel mindkét lehetőség fennáll.

A fenti eljárások megválasztását és hatékonyságát döntően befolyásolja a dokumentum-jellemzés teljessége. Ez a probléma éppen az automatikus indexelésnél jelentős, bizonyos tapasztalatok ugyanis arra mutatnak, hogy a teljes referátum-szöveg tárolása "tulzott teljességhez" vezethet, amelyet a kérdés megfelelő kezelésével kompenzálni kell.

A kifejezések súlyozása a manuális indexelésben rendszerint intuitív alapon történik, a kérdés feldolgozásánál. Az automatizálás lehetősége itt is statisztikai alapú. A súlyozás értékével kapcsolatban több kísérlet is volt, amelyek arra mutatnak, hogy hatékony súlyozáshoz meglehetősen komplikált, a gyűjtemény egészén alapuló eljárások szükségesek. A tapasztalatok mindenesetre azt jelzik, hogy a súlyozás az egyik, jellegzetesen automatizálandó eljárás az indexelésen belül.

Az indexnyelvnek a visszakeresésben elfoglalt helye ma még nem világos. A pontosság kedvéért meg kell különböztetni az index szótárat a különféle viszony- vagy

osztály-szerkezettől, amely a nyelvhez tartozik: Az utóbbi időben a legnagyobb vitát a nyelvbe beépülő ellenőrzés foka váltotta ki. Az automatizálás szempontjából a legfontosabb kérdések: kell-e szigorú ellenőrzés és ha igen, biztosítható-e gépi uton. Az eddigi tapasztalatok arra mutatnak, hogy egy szigorúan ellenőrzött nyelv gondosan megválasztott szótárral és kidolgozott viszony-szerkezettel nem fizetődik ki. A lényegi kérdés tehát az ellenőrzöttség foka.

A kulcsszavak osztályozásán alapuló rendszereknél a szótár-ellenőrzés két szinten lehetséges: a kulcsszavak kiválogatásánál vagy az osztálydeszkriptorok alkalmazásánál. Ez utóbbi azonban a nyelvi strukturához tartozik.

Nyilvánvaló, hogy a szótár-ellenőrzés statisztikai jellegű. Olyan szótár kell, amely a túlzottan gyakori vagy túlzottan ritka szavak használatát korlátozza. A legnagyobb idevágó kísérlet, Dennis-é ad hoc küszöbökkel operált, s visszakeresési eredményei gyengék voltak. További problémát jelent változó gyűjtemény esetén az indexszótár időnkénti felújítása. Az eddigi kísérletek alapján leszűrhető, hogy elsősorban a közepes gyakoriságú és viszonylag kevés együttes előfordulást mutató szavak érdekesek.

Az előbbtől eltérő stratégia az index-szótár súlyozással történő ellenőrzése. E súlyok ugyanis a gyűjtemény változása során könnyebben módosíthatók. Ez az eljárás nem zárja szükségképpen a gyakori kifejezéseket (mivel a kérdések igen gyakran tartalmaznak ilyeneket) és így javítja a visszahívást. A ritka kifejezések értéke még nem világos.

Az indexnyelven belüli osztályozás manuálisan a kifejezések tartalmán alapul. A gépi megoldás ehelyett csak a kifejezés "viselkedésén", a szövegen belüli eloszlásán alapulhat. A gépi indexelésnek az osztályozás talán legnehezebb pontja, annál is inkább, mert még a manuális módszerek sem mindenütt problémamentesek. Az eddigi kísérletek arra mutatnak, hogy lényegében elegendő a kifejezések közti direkt, "elsőrendű" kapcsolatok figyelembe vétele. A szükséges asszociációs mértékeket igen sokan tanulmányozták, és arra jutottak, hogy viszonylag egyszerű mérőszámok megfelelnek. Problémát okoznak viszont a nem-szignifikáns asszociációk, amelyeket ritka és gyakori kifejezések közt igen gyakran észleltek.

Az osztály-meghatározás igen fontos kérdésében már lényegesen kevesebb alkalmazható eredmény született. Ennek egyik oka, hogy a nem-hierarchikus osztályozás elméleti megalapozása hiányzik.

Az automatikus osztályozásra vonatkozó érdemi kísérletek csak 1965 után kezdődtek. Az erősen leegyszerűsített, asszociációs jegyzékeken alapuló eljárások nem vezettek megnyugtató eredményekre s egymásnak is ellentmondanak.

A teljes osztályozásra csupán három kísérlet zajlott le, általában gyenge, de legalábbis a gazdaságosságot megkérdőjelező eredményekkel. Egy működő rendszer alkalmaz automatikus osztályozást, de erősen leszűkített körben. Történt néhány kezdeményezés automatikus osztályozás alkalmazására teaurusz-szerkesztés során, ezek eredménye azonban még nem értékelhető.

A visszakeresés során a legfontosabb szemantikai eszközök a csonkolás és a súlyozás. A csonkolást számos működő rendszer haszonnal alkalmazza. A súlyozás hasznossága nem ennyire egyértelmű.

Az eddig tárgyalt direkt (a dokumentum nyelvén alapuló) indexelés mellett ujabban egyre nagyobb figyelemmel fordulnak a kutatók az indirekt indexelés felé, amelynek két legjellemzőbb formája a hivatkozások indexelési alkalmazása ill. a direkt dokumentum-osztályozás, a clusterképzés alkalmazása.

A hivatkozások visszakeresési alkalmazására viszonylag kevés kísérletet végeztek, mégis kellőképp megalapozottnak tűnik. Alkalmazták a hivatkozásokat cluster-képzésre is. Ennek ellenére az eddigi eredmények nem látszanak sem bizonyítani, sem cáfolni automatikus visszakeresésben a hivatkozások előnyét más módszerekkel szemben.

A dokumentum-clusterek képzését elsősorban a gépi keresési idő csökkentése indokolja, alkalmazása tehát csak nagy file esetén indokolt. Az eredményességet két szempont befolyásolja: elegendően közel állnak-e a dokumentum-clusterek a relevancia-clusterekhez ill. a clusterek mérete nem befolyásolja-e hátrányosan a visszahívást vagy a pontosságot. A tapasztalat ugyanis az, hogy a túl nagy cluster rontja a pontosságot, a túl kicsi a visszahívást. A hasonló dokumentumokból kialakított clusterok gazdaságossági előny mellett javíthatják a rendszer teljesítményét is.

Az eddig elvégzett cluster-képzési kísérletek két, bizonyos mértékig eltérő technikát alkalmaznak. Az egyik egy szinten választja szét a dokumentumokat, a mások többszintű hierarchiát alkalmaz. Az első esetben a hasonló dokumentumok egy mag körül tömörülnek és a mag leírása reprezentálja a clustert. A cluster meghatározása az eljárás többszöri iterációját igényli. A több-szintű cluster-képzésnél a hasonlóság fokának megfelelően a dokumentumok egyre bővebb csoportokat alkotnak. Mindkét eljárásra számos kísérletet folytattak le, eredményeik azonban még nem adnak világos képet az eljárás jellemzőiről.

GÉPESÍTETT RENDSZEREK

Eddig az automatikus indexelésről, mint önálló fejezetről volt szó. Ez a tevékenység azonban megfelelő gépesített rendszerek részeként kell működjék.

A működő gépi rendszerek gépesítettségük fokával és indexelési rendszerük bonyolultságával jellemezhetők. E rendszerek teljes áttekintésére ezuttal nem kerülhet sor, inkább bizonyos tipikus megközelítést említhetünk meg. A legalacsonyabb fokú gépesítésnek a gépi file-ban történő keresés tekinthető, ahol az indexelés és a kérdések feldolgozása manuálisan történik (INSPEC, Medlars stb.). Szempontunkból ezek a rendszerek érdektelenek.

Teljes, az elemzésre, leírásra és nyelv-előállításra kitérő automatizált rendszer általában igen egyszerű indexelést alkalmaz. Rendszerint a címből, egyes szavakból vagy szó-sorozatokból keresnek vissza. A cím - elsősorban gazdasági okokból - igen népszerű. A teljes dokumentum-szöveg alkalmazása igen ritka.

Ugy tűnik, az egyetlen, teljesen automatizált működő rendszer, amely bonyolult indexelést alkalmaz, Hillman Leadermart rendszere. Ez vegyes adatbázisú on-line rendszer.

Az előbbi, standard rendszerek mellett működik számos, a használó közreműködésére is építő nem-standard rendszer is. Közülük elsősorban az interaktív on-line

visszakeresés vált népszerűvé, szinte divatossá. A használói részvétel foka ezekben igen változó: néhol nincs is szó valódi részvételről, másutt a kérdés újra-fogalmazására van lehetőség vagy csak a relevancia-visszacsatolás történik interaktív módon.

Az automatikus indexelés szempontjából az interaktív keresés elsősorban azért érdekes, mert a használó beavatkozási lehetősége esetleg enyhíti a dokumentum indexelésével szemben támasztott követelményeket, illetve a keresési visszacsatolás helyes irányban módosíthatja a dokumentum leírását. Az eddigi tapasztalatok azonban nem látszanak elegendőnek ahhoz, hogy az interaktív keresésnek a gépi indexelésre kifejtett hatását véglegesen meghatározottnak tekinthessük.

A másik terület, ahol a használók közreműködése jelentőséggel bír, a nyomtatott index, s az itt szerzett tapasztalatok egy része felhasználhatónak látszik az automatikus indexelés továbbfejlesztésében is.

ÉRTÉKELÉSI KÍSÉRLETEK

Az egyes részeredmények értékelésére eddig tett kísérletek mellett szólni kell néhány szót az automatikus indexelés, mint egész értékeléséről is. Kényelmi szempontból célszerű ezeket a rendszereket két csoportra osztani: egyszerű és összetett indexelési rendszerekre. Az elsőre példa a cím használata vagy a csonkolás, a másodikra az input szöveg elemzése, a tezaurusz deskriptorok alkalmazása, a tartalmi viszony-jelölők vagy az osztályozott struktúra a visszakeresésben.

Salton már 1970-ben igazolta, hogy az egyszerűbb gépi eljárások versenyképesek a bonyolult manuális eljárásokkal. A teljes, belső és külső értékelés eredménye tulajdonképpen páronkénti összehasonlítások sorozataként fejezhető ki, amelyeket külön-külön is igazolni kellene. Egy ilyen sorozat:

összetett kézi indexelés	jobb mint	egyszerű kézi indexelés
összetett gépi indexelés	jobb mint	egyszerű gépi indexelés
egyszerű gépi indexelés	jobb mint	egyszerű kézi indexelés
összetett gépi indexelés	jobb mint	összetett kézi indexelés
egyszerű gépi indexelés	jobb mint	összetett kézi indexelés
összetett gépi indexelés	jobb mint	egyszerű kézi indexelés

Az eddigi, igen nagy számu, de egymással sokszor összehasonlithatatlan értékelési kísérlet alapján a következő megállapítások látszanak leszűrhetőnek:

- az összetett kézi indexelés lényegesen jobb eredményeket biztosít, mint az egyszerű kézi indexelés;

- az összetett gépi indexelés általában nem biztosít sem jobb, sem rosszabb eredményeket, mint az egyszerű gépi indexelés (ez utóbbihoz azonban két kiegészítő megállapítás járul: a referátumból kiinduló egyszerű gépi indexelés lényegesen jobb, mint a címre szorító egyszerű gépi indexelés, ill. a gyűjteményi gyakoriságon alapuló súlyozás hozzájárul a visszakeresés javításához);

- az egyszerű gépi és egyszerű kézi indexelés összehasonlítását célzó kísérletek csak azt bizonyították, hogy a kulcsszavak alkalmazása eredményesebb, mint akár a címek, akár a referátumok felhasználása;
- az egyszerű gépi ill. az összetett kézi indexelés összevetéséből az összetett kézi indexelés látszik hatékonyabbnak (bár sok esetben a teljesítmény igen hasonló);
- az összetett gépi indexelés hatékonysága lényegében megegyezik az egyszerű kézi indexeléssel;
- az összetett kézi indexelés lényegesen jobb, mint az összetett gépi indexelés.

KÖVETKEZTETÉSEK

Bár az éppen most összefoglalt eredmények igen eltérő vizsgálatokból származnak, s így egyes részletekben a későbbi kísérletek azokat módosíthatják, az automatikus indexelés jelenlegi helyzetéről mégis adnak egy átfogó képet.

A vonatkozó kutatásokra elsősorban azok alacsony száma jellemző. Ennek megfelelően működő rendszereken hatásuk szinte észrevehetetlen.

A kutatások és kísérletek kevés véglegesnek tekinthető eredményének egyike, hogy bizonyos egyszerű módszerek, pl. kulcsszavak használata vagy statisztikai alapú sulyozás, igen jól használhatók. Ennek megfelelően a következő működő gépi rendszer látszik elképzelhetőnek: Inputként cím+referátum, elemzés negatív szótárral, irreleváns szavak törlése; leírás során gyakorisági sulyozás; az index nyelvben szóvégződések automatikus levágása; keresésben gyűjteményi gyakorisági sulyozás, output rendezése normalizált összehasonlító függvényvel. Ez a felvázolt lehetőség nem tulságosan izgalmas és véglegesnek semmiképp sem tekinthető. Elsősorban és mindenekelőtt további kutatásokra van szükség ahhoz, hogy a téma pontos képét és lehetőségeit lássuk.