ЕВРОПЕЙСКИ СЪЮЗ
Европейски фонд
за регионално развитие
Инвестираме във вашето бъдеще

ОПЕРАТИВНА ПРОГРАМА
„Развитие на конкурентоспособността
на българската икономика" 2007-2013
www.opcompetitiveness.bg

ПРОЕКТ: Създаване на алгоритми за извличане на критерии за сортиране, групиране и управление на документи в големи масиви от структурирана информация чрез интеграция на международни класификации и резултати от пълнотекстово търсене в многоезичен режим

Обща информация

С този проект за научно изследване Ламел ООД разработва алгоритми за динамично извличане на критерии за търсене в големи масиви от структурирани документи които да интегрират международни класификации на няколко езика и резултати от пълнотекстово търсене.

Те ще могат да бъдат внедрени във всеки интерфейс за боравене с информация или машина за търсене, включително интернет търсачки, библиотечни системи боравещи с пълния текст на печатните издания, патентни бази данни и други подобни. Целта на алгоритмите е да позволяват на потребител, незапознат с цялостната структура на класификаторите, да намали броя на операциите за намиране на необходимите му документи на базата на стъпки, дефиниращи приликите и отличията между избраните на всеки етап от търсенето документи, както и да направи възможно търсенето и намирането на релевантни документи на всички достъпни езици, без да е необходимо потребителят да ги познава.

Дейностите по проекта включват проектирането на структури за въвеждане и бази данни от структурирани документи, многоезикови лексикални съответствия към многокритериални класификации и лексикални спектри, разработването на същинските алгоритми за извличане на критерии за търсене от лексикалните спектри и изследване на необходимия за надеждното прилагате на алгоритмите изчислителен ресурс при различни работни натоварвания.

За нуждите на алгоритмите, пълният текст на всеки един документ в базата данни трябва да бъде анализиран с числови и статистически методи за определяне на значещите думи в текста и тяхната честота на появяване, като тези данни съставят така наречените от нас лексикални спектри на документа. Статичният лексикален спектър ще показва степента на съответствие на документа към всяко ниво на съществуващите класификации, а динамичните лексикални спектри ще показват това съответствие в сравнение с останалите документи от дефинирана група. В резултат на това динамичните лексикални спектри могат да осигурят подаването на относително малък брой отличия в произволно избрана група документи, което да доведе до създаване на практически интерфейси, позволяващи на потребителя да намира много по-бързо

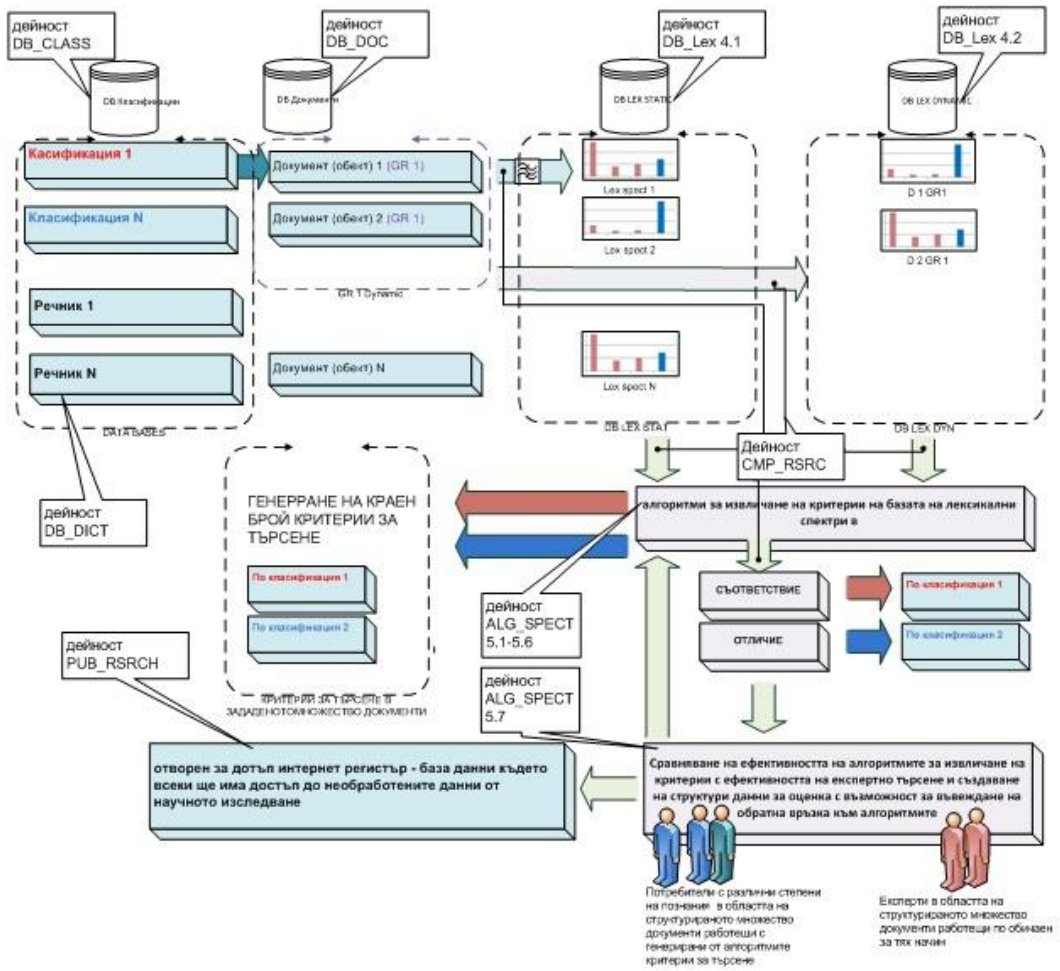
Този документ е създаден по проект: BG161PO003-1.1.06-0088-C0001 Изследване на функционален модел на ново поколение компютърни интерфейси за сортиране, групиране и управление на информационни обекти на компютърно устройство
Бенефициент: **Ламел ООД**
Документът е създаден с финансовата подкрепа на Оперативна програма „Развитие на конкурентоспособността на българската икономика" 2007-2013, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие. Цялата отговорност за съдържанието на документа се носи от Ламел ООД и при никакви обстоятелства не може да се приема, че този документ отразява официалното становище на Европейския съюз и Договарящия орган.

ЕВРОПЕЙСКИ СЪЮЗ
Европейски фонд
за регионално развитие
Инвестираме във вашето бъдеще

ОПЕРАТИВНА ПРОГРАМА
„Развитие на конкурентоспособността
на българската икономика" 2007-2013
www.opcompetitiveness.bg

релевантна информация в големи масиви със сходни документи по пътя на сравнението и изключването.

Базите данни на класификаторите ще описват съществуващи международно признати и използвани класификации от различен тип като за целта на разработката ще използваме Международната патентна класификация в комбинация с една или няколко други международно признати многоезични класификации като например Международната класификация на стоките и услугите за регистрация на марки, Женевска класификация на химичните вещества, класификации по Евростат или митнически класификации. Успешното структуриране на бази данни за няколко от тези класификатори би трябвало да доведе до извеждането на удобна система за последващо въвеждане на всякакви специфични системи за класификация или елементи от системи за класификация, които потребителят би могъл да ползва.



Този документ е създаден по проект: BG161PO003-1.1.06-0088-C0001 Изследване на функционален модел на ново поколение компютърни интерфейси за сортиране, групиране и управление на информационни обекти на компютърно устройство
Бенефициент: **Ламел ООД**
Документът е създаден с финансовата подкрепа на Оперативна програма „Развитие на конкурентоспособността на българската икономика" 2007-2013, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие. Цялата отговорност за съдържанието на документа се носи от Ламел ООД и при никакви обстоятелства не може да се приема, че този документ отразява официалното становище на Европейския съюз и Договарящия орган.

За ефективността на алгоритмите ще се съди от количественото сравнение между получените параметри с изведени от експерт в областта на изследваните документи. За целта ще се разработят средства за въвеждане на стойностите в структурата данни на модела, които ще могат да въведат обратна връзка в алгоритмите, която да прецизира резултатите на получаваните критерии за търсене сортиране и групиране както на базата на количествени (статистически), така и на базата на качествени (експертна оценка) критерии. Заложените в методологията оценки на входните параметри трябва по структура да отговарят на международните стандарти за описание на данните като в случая е заложен стандартът на Евростат. Този подход е ключов за оценка на релевантността на изходните данни при използване на специализирани класификации въвеждани при бъдещо практическо използване на метода.

Продължителността на проекта е 24 месеца като за това време трябва да бъдат изведени алгоритми за извличане на критерии за сортиране, групиране и управление на документи в големи масиви от структурирана информация, описани в подробен доклад, утвърдена структура за 4 бази данни, както са описани в дейностите, с възможност за въвеждане на допълнителна информация и различни по структура класификатори, подробен окончателен доклад за необходимия за приложението на алгоритмите изчислителен ресурс при различни условия и за различни устройства и отворен интернет регистър, където всеки ще има достъп до необработените данни от научното изследване.

Този документ е създаден по проект: BG161PO003-1.1.06-0088-C0001 Изследване на функционален модел на ново поколение компютърни интерфейси за сортиране, групиране и управление на информационни обекти на компютърно устройство

Бенефициент: **Ламел ООД**

Документът е създаден с финансовата подкрепа на Оперативна програма „Развитие на конкурентоспособността на българската икономика" 2007-2013, съфинансирана от Европейския съюз чрез Европейския фонд за регионално развитие. Цялата отговорност за съдържанието на документа се носи от Ламел ООД и при никакви обстоятелства не може да се приема, че този документ отразява официалното становище на Европейския съюз и Договарящия орган.