



ЕВРОПЕЙСКИ СЪЮЗ  
Европейски фонд  
за регионално развитие  
*Инвестираме във вашето бъдеще*



[www.eufunds.bg](http://www.eufunds.bg)



ОПЕРАТИВНА ПРОГРАМА  
„Развитие на конкурентоспособността  
на българската икономика” 2007-2013  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

## ТЕХНИЧЕСКИ СПЕЦИФИКАЦИИ И ПЪЛНО ОПИСАНИЕ НА ОБЕКТА НА ПРОЦЕДУРАТА

### 1. Обект на процедурата:

Обектът на настоящата процедура за избор на изпълнител е:

**Доставка и въвеждане в експлоатация на хардуер и софтуер по обособени позиции:**

#### **Обособена позиция 1 - Хардуер:**

- Сървър – 2 броя;
- Работна станция – 5 броя;
- Работна станция iMac (или еквивалент) – 1 брой;
- Тестово мобилно устройство – 1 брой;
- Тестово мобилно устройство – 1 брой;
- Тестово мобилно устройство – 1 брой.

#### **Обособена позиция 2 - Софтуер:**

- Специализиран софтуер - система за компютърна обработка на естествен език (NLP) – 1 брой;
- Специализиран софтуер за автоматична категоризация на български език – 1 брой.

2. Срок на изпълнение: Срокът за изпълнение е **14 календарни дни** (от сключване на договора за доставка), но в рамките на договора за безвъзмездна финансова помощ

### 3. Технически спецификации:

#### ОПИСАНИЕ:

##### Обособена позиция 1 - Хардуер:

- Сървър – 2 броя:

**Минимални технически и функционални характеристики за Сървър – 2 броя:**

Процесор: не по-малко от 4 ядрен; не по-малко от 2 броя процесори за всеки сървър; не по-малко от 2.4 GHz;

Памет: не по-малко от 24GB;

Твърд диск: не по – малко от 6x300 GB;

Наличие на DVD – RW оптично устройство;

Мрежа: 1 GbE; 2 порта;

Наличие на захранване.

**Допълнителни технически и функционални характеристики, предмет на оценка за Сървър – 2 броя:**

*Съгласно утвърдената методика, ще бъдат оценявани:*

Процесор

Памет

Твърд диск

Наличие на интерфейс за отдалечено управление и наблюдение на сървърите

Възможност за разширение

- Работна станция – 5 броя:

**Минимални технически и функционални характеристики за Работна станция – 5 броя:**

Тип: лаптоп;

*Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтуер” ЕООД.*

*Този документ 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 „Атлас Софтуер” ЕООД 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*



[www.eufunds.bg](http://www.eufunds.bg)



**ОПЕРАТИВНА ПРОГРАМА**  
**„Развитие на конкурентоспособността**  
**на българската икономика” 2007-2013**  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

Наличие на Bluetooth;

Наличие на безжична мрежа Wi – Fi;

Процесор: не по-малко от 4 ядрен; не по-малко от 2.2 GHz;

Памет: не по-малко от 8GB

Твърд диск: не по – малко от 750 GB;

Дисплей: не по – малко от 17”;

Батерия: не по-малко от 6 клетъчна; литиево – йонна;

Наличие на DVD – RW оптично устройство;

Наличие на захранване.

**Допълнителни технически и функционални характеристики, предмет на оценка за Работна станция – 5 броя:**

*Съгласно утвърдената методика, ще бъдат оценявани:*

Процесор

Резолуция на дисплея

Памет

Твърд диск

- **Работна станция iMac (или еквивалент) – 1 брой**

**Минимални технически и функционални характеристики за Работна станция iMac (или еквивалент) – 1 брой:**

Тип: всичко в едно;

Процесор: не по-малко от 4 ядрен; не по-малко от 3.1 GHz;

Дисплей: не по-малко от 27”;

Наличие на захранване;

Наличие на мрежова карта;

Памет: не по-малко от 4 GB;

Твърд диск: не по-малко от 1 TB

Видео карта: не по-малко от 1 GB.

**Допълнителни технически и функционални характеристики, предмет на оценка за Работна станция iMac (или еквивалент) – 1 брой:**

*Съгласно утвърдената методика, ще бъде оценявана:*

Резолуция на дисплея

- **Тестово мобилно устройство – 1 брой**

**Минимални технически и функционални характеристики за Тестово мобилно устройство – 1 брой:**

Тип: смартфон;

Процесор: на по-малко от 2 ядрен; не по-малко от 1 GHz;

Възможност за работа под операционна система iOS;

**Допълнителни технически и функционални характеристики, предмет на оценка за Тестово мобилно устройство – 1 брой:**

*Съгласно утвърдената методика, ще бъде оценявана:*

Резолуция на дисплея

- **Тестово мобилно устройство – 1 брой**

**Минимални технически и функционални характеристики за Тестово мобилно устройство – 1 брой:**

Тип: таблет;

Процесор: на по-малко от 2 ядрен; не по-малко от 1 GHz;

Наличие на безжична мрежа;

Дисково пространство: не по-малко от 16 GB;

Размер на дисплея: не по-малко от 9.7”;

Възможност за работа под операционна система iOS;

**Допълнителни технически и функционални характеристики, предмет на оценка за Тестово мобилно устройство – 1 брой:**

*Съгласно утвърдената методика, ще бъдат оценявани:*

Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтнер” ЕООД.

Този документ 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 „Атлас Софтнер” ЕООД 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*



[www.eufunds.bg](http://www.eufunds.bg)



ОПЕРАТИВНА ПРОГРАМА  
„Развитие на конкурентоспособността  
на българската икономика” 2007-2013  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

Резолюция на дисплея

3 G функция

- **Тестово мобилно устройство – 1 брой**

**Минимални технически и функционални характеристики за Тестово мобилно устройство – 1 брой:**

Тип: таблет;

Процесор: на по-малко от 2 ядрен; не по-малко от 1,4 GHz;

Наличие на безжична мрежа;

Дисково пространство: не по-малко от 16 GB;

Размер на дисплея: не по-малко от 4”;

Възможност за работа под операционна система Android;

**Допълнителни технически и функционални характеристики, предмет на оценка за за Тестово мобилно устройство – 1 брой:**

*Съгласно утвърдената методика, ще бъдат оценявани:*

Резолюция на дисплея

3 G функция

**Важно: Допълнителните технически и функционални характеристики на хардуера са предмет на оценка и нямат задължителен характер. Офертата на кандидат, предложил хардуер без наличие на съответните допълнителни параметри, не може да бъде отхвърлена на това основание.**

## Обособена позиция 2 - Софтуер:

**Специализиран софтуер - система за компютърна обработка на естествен език (NLP) – 1 брой;**

**Минимални технически и функционални характеристики за система за компютърна обработка на естествен език (NLP) – 1 брой:**

Сегментатори:

1. разделител на параграфите в текста - компонентът аотира начало и края на всеки параграф в текста.
2. разделител на изреченията във всеки параграф от текста - компонентът аотира началото и края на всяко изречение във всеки параграф в текста. Допълнително, компонентът трябва да разпознава и отбелязва дали изречението покрива минимални изисквания за граматическа вярност (например, да има поне един глагол) и дали изречението е част от пряка или не-пряка реч.
3. разделител на “думите” във всяко изречение от текста - тук “дума” се използва в широк смисъл - освен стандартното значение на “дума”, компонентът трябва да аотира числа, пунктуация, чужди думи и т.н.

Анализатори: Следните компоненти са необходими, за семантичен анализ на входния текст:

1. Отбелязване на частите на речта - този компонент трябва да отбележи като каква част от речта (съществително, прилагателно, глагол, местоимение и т.н.) участва всяка дума от текста. Описание на разпознаваните части от речта трябва да бъдат неделима част от документацията на доставения софтуер.
2. Определяне на основната форма на думата - този компонент трябва да отбележи основната форма на всяка дума от текста.

Анотатори на по-сложни изрази: Компонентите в тази секция трябва да аотират изрази, съставени от повече от една дума или изпълняващи специфична функция в текста.

1. Отбелязване на фрази от групата на съществителното - компонентът трябва да отбелязва прости фрази, напр. “синьо небе”, “хубава жена”, както и съставни фрази, например, “объвките на хубавата жена”. Анотацията на всяка фраза трябва да съдържа и началото и края на основното съществително във фразата.
2. Отбелязване на именни фрази - компонентът трябва да отбелязва именана хора (българи и чужденци), географски места, организации и фирми, дати, проценти, адреси в Интернет (URL), електронни пощи, които се срещат в изходния текст. Анотацията на всяка именна фраза трябва да съдържа типа на разпознатото име.

Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтуер” ЕООД.

Този документ 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 „Атлас Софтуер” ЕООД 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*



[www.eufunds.bg](http://www.eufunds.bg)



**ОПЕРАТИВНА ПРОГРАМА**  
„Развитие на конкурентоспособността  
на българската икономика” 2007-2013  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

Други:

1. определяне значението на думите в текста (т.нар. Word Sense Disambiguation) - компонентът трябва с голяма степен на точност да определи кои смисъл на дадена дума е използван в контекста на изречението. Напр. в следните изреченията, думата “гума” се използва по различен начин, определен от смисъла:

а. “Гумата на колата се спука”;

б. “Иван изтри написаното с гума”.

2. достъп до WordNet на български - ще бъдат разработени иновативни компоненти, които използват български WordNet за по-точен семантичен анализ на текста. Наличието на български WordNet или компонент, които да осъществява достъп до български WordNet е предимство.

Изисквания за интеграция на компонентите

1. Основната платформа за разработване на “Документед.бг” ще бъде базирана на JAVA 6. Компонентите, изброени в първата секция на документа трябва да могат да се интегрират в JAVA-базирано приложение.
2. Всяка оферта, трябва да предложи архитектура за интегрирането на компонентите в приложение, което получава входен текст, а очакваният изход са всички анотации, които са предоставени от компонентите. Предимство е използването на стандарни и/или добре известни подходи и архитектури за интегриране на хетерогенни компоненти за автоматичен анализ на естествени езици.
3. Предложената архитектура или интеграционен подход трябва да позволява лесно и бързо хоризонтално разширяване (horizontal scalability).
4. Всеки от компонентите, както и цялостно-изграденият софтуер трябва да предлагат оптимални стойности на прецизност съпоставени към времето, необходимо за намирането на анотациите във входния текст. Всяка оферта трябва да представи стойности за прецизността на всеки от компонентите, както и времето, което всеки от компонентите използва за обработка на текст с не повече от 1'000, 2'000, 4'000, 8'000, 16'000, 32'000 и 64'000 “думи”.
5. Забележка: компонентите, описани в първата част на това задание могат да се комбинират или композират, ако използват единни и същи ресурси и времето за работа на мета-компонентът е по-добро от сбора от времената на композираният компоненти.

**Допълнителни технически и функционални характеристики, предмет на оценка за система за компютърна обработка на естествен език (NLP) – 1 брой:**

Съгласно утвърдената методика, ще бъдат оценявани:

Допълнителна функционалност за грануларно определяне на подтипове на имената и за нормализиране на времевите изрази

Допълнителен компонент Сегментатор да разпознава отделните прости изречения (clauses) в по-голямо изречение

Допълнителен компонент Анализатор да отбелязва съществени характеристики на формите на думите: род, число, определеност, лице, време, както и характеристики, които са важни при семантичен анализ: съществително - нарицателно и собствено, съществително - лице, одушевено и неодушевено, глагол — личен и безличен, глагол — преходен и непреходен

Допълнителен компонент Анотатор на сложни фрази да отбелязва фрази и имена с различна дължина и съответно да може да бъде настройван за това

**Специализиран софтуер за автоматична категоризация на български език – 1 брой.**

**Минимални технически и функционални характеристики за софтуер за автоматична категоризация на български език – 1 брой:**

Функционални изисквания

Софтуерният модул за АКТД трябва да изпълнява следните задачи:

1. Да създава категоризационни модели на база множество от категории (етикети) и принадлежност на текстови документи за трениране към тези категории. **Забележка 1:** “думите” във входните документи са представени като числа, за да се избегне повторна и ненужна обработка на естествения език на текста. АКТД няма нужда да използва средства за обработка на естествен език, за това АКТД по същество е езиково-независим софтуерен

Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтуер” ЕООД.

Този документ 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 „Атлас Софтуер” ЕООД 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*



[www.eufunds.bg](http://www.eufunds.bg)



ОПЕРАТИВНА ПРОГРАМА  
„Развитие на конкурентоспособността  
на българската икономика” 2007-2013  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

компонент:

Забележка 2: в общия случай, релацията между категория и документ е “много-към-много” и за това изражданите модели ще се използват за решаване на “multi-class multi-label” категоризационна задача;

2. Да оценява категоризационен модел на база множество от тестови документи, предварително категоризирани с етикетите от оценявания модел;
3. Да намира оптимален модел чрез итерации на задачи 1 и 2 с цел да се постигне максимална F1 оценка и минимален (като размер) модел.
4. Да намира най-подходящите N категории за непознат документ (документ, който не е използван за трениране на моделите), като за една от тези N категории се връща и степен на увереност (prediction confidence).
5. Да предоставя възможност за групиране на подобни документи, като за всяка група се връщат и общите характеристики на документите в групата.

При израждането на категоризационните модели и решаването на задачи 4 и 5 трябва да се обединят резултатите от няколко категоризационни алгоритми, по възможност, базирани на различни парадигми: категоризиране по съседство, чрез дървета за избор (decision trees), базирани на теорията на вероятностите; logistic regression и SVM (Support Vector Machines). Оферти, които използват повече категоризационни алгоритми, както и утвърдени методи за обединяване на резултатите ще имат предимство.

Не-функционални изисквания

1. Основната платформа за разработване на “Документед.бг” ще бъде базирана на JAVA 6. Компонентите, изброени в първата секция на документа трябва да могат да се интегрират в JAVA-базирано приложение.
2. Софтуерният модул трябва да може да работи ефективно с не по-малко от 500 категории и не по-малко от 100'000 документи за трениране на модели, използвайки не повече от 8Gb JAVA heap memory.
3. Всяка оферта, трябва да предложи архитектура за интегрирането на компонентите в приложение. Предимство е използването на стандартни и/или добре известни подходи и архитектури за интегриране на хетерогенни компоненти и алгоритми за автоматична категоризация на текстови документи.
4. Предложената архитектура или подход за интеграция трябва да позволява лесно и бързо хоризонтално разширяване (horizontal scalability).
5. Всеки от компонентите, както и цялостно-израденият софтуер трябва да предлагат оптимални стойности на precision и recall, съпоставени към времето, необходимо за решаване на задачите, описани в “Функционални изисквания”. Всяка оферта трябва да представи стойности за precision и recall както и времето, за изпълнение на задачите от 1 до 5 върху тестови данни, предоставени от възложителя на настоящото задание.

**Допълнителни технически и функционални характеристики, предмет на оценка за софтуер за автоматична категоризация на български език – 1 брой:**

*Съгласно утвърдената методика, ще бъдат оценявани:*

Наличие и на 4-те основни алгоритми за категоризиране на текстови документи – SVM (или еквивалент),

Naive Bayesian (или еквивалент), Relative Entropy (или еквивалент), CFC (или еквивалент)

Наличие на допълнителни алгоритми за категоризация

Възможност за работа на софтуерния пакет върху стандартен (не-кълъстиран) компютърен хардуер

**Важно: Допълнителните технически и функционални характеристики на софтуера са предмет на оценка и нямат задължителен характер. Офертата на кандидат, предложил софтуер без наличие на съответните допълнителни параметри, не може да бъде отхвърлена на това основание.**

**В доставката на оборудването се включва още:**

- Доставка и въвеждане в експлоатация:

Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтуер” ЕООД.

Този документ 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 „Атлас Софтуер” ЕООД 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*



[www.eufunds.bg](http://www.eufunds.bg)



ОПЕРАТИВНА ПРОГРАМА  
„Развитие на конкурентоспособността  
на българската икономика” 2007-2013  
[www.opcompetitiveness.bg](http://www.opcompetitiveness.bg)

Доставката и въвеждането в експлоатация на хардуера и софтуера е за сметка доставчика и се извършва в офиса на „Атлас Софтуер“ ЕООД на адрес: гр. София, ул. Професор Николай Генчев, №18.

• **Изисквания към Гаранция и след сервизна поддръжка – САМО ЗА ОБОСОБЕНА ПОЗИЦИЯ 1:**

Кандидатите по ОБОСОБЕНА ПОЗИЦИЯ 1 трябва да са в състояние да предоставят минимум 12 месеца гаранция и сервизна поддръжка на доставения хардуер. Гаранцията и сервизната поддръжка започват да действат от момента на финално приемане на хардуера;

• **Ограничения при изпълнение на поръчката:**

- Кандидатите имат право да представят само една оферта (техническа и ценова) за участие в процедурата;
- Кандидат в процедурата може да бъде всяко българско или чуждестранно физическо или юридическо лице или техни обединения;

*Този документ е създаден по проект BG161PO003-1.1.05-0165-C0001 „Разработване на иновативна онлайн услуга „Documented.bg”, Бенефициент: „Атлас Софтуер” ЕООД.*

*Този документ 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 „Атлас Софтуер” ЕООД и при никакви обстоятелства не може да се приема, че този документ отразява официалното становище на Европейския съюз и Договарящия орган.*