



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ



BG16RFOP002-1.005-0232-C01 „STARE – СИСТЕМА ЗА ИДЕНТИФИКАЦИЯ И МОНИТОРИНГ НА СВЪРЗАНИ С ДЪРЖАВАТА ПРЕДПРИЯТИЯ”

ТЕХНИЧЕСКА СПЕЦИФИКАЦИЯ

I. ПРЕДМЕТ НА ПУБЛИЧНАТА ПОКАНА

Предметът на настоящата процедура за избор на изпълнител с публична покана е: „Закупуване на корпус с минимум 50 000 статии за трениране на семантични технологии”.

II. ОБХВАТ И ОПИСАНИЕ НА ДОСТАВКАТА

„А ДЕЙТА ПРО“ ООД е бенефициент по Договор за безвъзмездна финансова помощ (ДФП) BG16RFOP002-1.005-0232-C01 за проект „StaRe – система за идентификация и мониторинг на свързани с държавата предприятия”, финансиран по ОП “Иновации и конкурентоспособност” 2014-2020 със срок за изпълнение 25.02.2020 г.

Проектната иновация StaRe (от англ. наблюдавам) цели проектиране и разработка на нов ИКТ продукт, обединяващ веб-базирани ресурси, методи за извличане на данни от тях и технологии за автоматична обработка. Предназначението му е предоставяне на единна, актуална, широкообхватна и структурирана информация за предприятията в света, които под една или друга форма, директно или индиректно (вкл. чрез физически лица) са обвързани с държавни и общински органи, служби и регулатори. Иновацията, предмет на разработка, е с приложимост на пазара на бизнес интеледжънс информационни продукти, бележещ многократен ръст в сегмента Governance, Risk+Compliance във връзка с регулациите за мерките срещу изпирането на пари и финансирането на тероризма (KYC/AML), въведени в развитите страни и формиращи значително търсене на бази данни с терористи и криминално проявени лица, данни за политически фигури, за санкционирани лица и организации, за компании, собственост на държавата, за санкционирани компании и др.

За реализирането на изследователските задачи в проекта е предвидено закупуване на материали – база данни с минимум 50 000 статии на различни езици в областта на управлението, риска и съответствието (Governance, Risk and Compliance – GRC). Чрез влагането и преработването на тези материали ще бъде осъществено кодирането и машинното обучение на семантичните алгоритми и системи за разпознаване на свързани с държавна намеса събития и организации.

----- www.eufunds.bg -----

BG16RFOP002-1.005-0232-C01 „STARE – СИСТЕМА ЗА ИДЕНТИФИКАЦИЯ И МОНИТОРИНГ НА СВЪРЗАНИ С ДЪРЖАВАТА ПРЕДПРИЯТИЯ”

1. МИНИМАЛНИ ИЗИСКВАНИЯ КЪМ БАЗАТА ДАННИ:

1.1. Вид на статиите, включени в базата данни:

Кандидатът следва да достави актуални новинарски статии (от вестници и списания). Нюзлетъри, брийфове, списъци, таблици не са приемливи поради нуждата от контекст за разглежданите субект и тема.

Статиите следва да са организирани в база данни с минимум следните полета:

- заглавие,
- дата на публикуване,
- източник,
- уникален идентификатор URL,
- тяло, изчистено от всякакъв код (html, javascript и пр.),
- пълно име на субектите - от съдържанието на всички налични текстове следва да са извлечени в отделни полета на базата данни засегнатите физически и юридически лица. Следва да са посочени пълните имена на субектите, като за физическо лице следва да се предоставят минимум лично име и фамилия.
- анотации по темите, посочени в т. 1.2 – да се посочат всички приложими теми (multi-label classification), съгласно таксономията от 1.2.

За гарантиране качеството на машинното обучение на семантичните алгоритми и системи, предоставената база данни следва да съдържа единствено статии, публикувани в рамките на 1 (една) година от датата на обявяване на настоящата процедура. Това се налага предвид скоростта на обновяване на съвременния език и необходимостта от актуалност на изразните средства, използвани в новинарските статии.

1.2. Насоченост на статиите, включени в базата данни

Статиите в базата данни следва да обхващат новини в областта на управлението, риска и съответствието (Governance, Risk and Compliance – GRC) с насоченост в следните теми:

1. Корпоративни финанси,
2. Правителствени договори,
3. Сливания и придобивания,
4. Приватизационни и национализационни сделки и промяна на собственост,
5. Политически, пазарни и борсови новини,
6. Рискови новини с престъпления.

BG16RFOP002-1.005-0232-C01 „STARE – СИСТЕМА ЗА ИДЕНТИФИКАЦИЯ И МОНИТОРИНГ НА СВЪРЗАНИ С ДЪРЖАВАТА ПРЕДПРИЯТИЯ”

1.3. Език на статиите, включени в базата данни

Базата данни следва да съдържа статии на минимум следните езици: английски, немски, френски, испански, китайски и турски език. Статиите от всеки един от посочените езици следва да съставляват поне 10% от общия брой статии, предоставени в базата данни.

1.4. Формат на базата данни

Базата данни следва да бъде представена във формат, лесен за конвертиране и автоматична обработка на данните (напр. XML или еквивалент).

2. УСЛОВИЯ ЗА ДОСТАВКА

2.1. Място и начин на доставка

В рамките на срока за доставка, Кандидатът предоставя базата данни:

- в електронен вариант на електронна поща, посочена от Възложителя, или
- на CD или флаш памет на адрес: гр. София, ул. „Кирил и Методий“ № 64.

2.2. Срок за доставка

Кандидатът следва да предложи срок (в календарни дни), в рамките на който ще достави базата данни. Срокът за доставка не може да е по-дълъг от 15 календарни дни считано от датата на сключване на договора за доставка.

Следва да се има предвид, че срокът за доставка не включва провеждането на едnodневно обучение по т. 3.

3. ОБУЧЕНИЕ ЗА РАБОТА С БАЗАТА ДАННИ

Кандидатът следва да проведе за собствена сметка едnodневно обучение за работа с базата данни на служители, посочени от Възложителя.

Обучението следва да се проведе в срок до 10 календарни дни считано от датата на доставка.

III. ИЗИСКВАНИЯ КЪМ ТЕХНИЧЕСКОТО ПРЕДЛОЖЕНИЕ

В техническото си предложение Кандидатът задължително посочва:

----- www.eufunds.bg -----



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ ФОНД ЗА
РЕГИОНАЛНО РАЗВИТИЕ



ОПЕРАТИВНА ПРОГРАМА
ИНОВАЦИИ И
КОНКУРЕНТОСПОСОБНОСТ

BG16RFOP002-1.005-0232-C01 „STARE – СИСТЕМА ЗА ИДЕНТИФИКАЦИЯ И МОНИТОРИНГ НА СВЪРЗАНИ С ДЪРЖАВАТА ПРЕДПРИЯТИЯ”

- Информация как предложената от него база данни отговаря на всяко от изисквания, посочените в П.1.;
- Информация за условия на доставката (място, начин и срок на доставка), съгласно посоченото в П.2.;
- Потвърждение за провеждане на обучение, съгласно посоченото в т. П.3.

IV. ОЧАКВАНИ РЕЗУЛТАТИ

1. Доставена база данни (корпус) със статии, напълно съответстваща на посочените минимални изисквания;
2. Проведено еднодневно обучение за работа с базата данни.

----- www.eufunds.bg -----