

Dyrektor Generalny
Magdalena Maciejewska

Warszawa, 7.10.2019 r.

Odpowiedzi na pytania
w konkursie pn. „System klasyfikowania zgłoszeń – GovTech Polska”
(nr postępowania: BG-II.211.19.2019)

Pytanie 1 Uczestnika konkursu:

Jaki okres czasu (co do daty publikacji dokumentu) będą obejmowały zgłoszenia w zbiorze testowym na Etapach I i II? Czy zbiór testowy będzie zawierał dokumenty archiwalne, o dacie np. sprzed roku 2019?

Odpowiedź Zamawiającego:

Zbiór testowy będzie zawierał zgłoszenia niepubliczne, które nie zostały jeszcze ujawnione (przed publikacją w Biuletynie Urzędu Patentowego i przed udostępnieniem w „Wyszukiwarce wynalazków i wzorów użytkowych – Register Plus” – dostępnej na stronie internetowej UPRP). Publikacja następuje po 18 miesiącach od daty zgłoszenia, zatem zbiór testowy (identyczny dla wszystkich zgłoszonych rozwiązań podlegających ocenie) w Etapie I będzie zawierał zgłoszenia wynalazków i wzorów użytkowych dokonane między majem 2018 r., a listopadem 2019 r. natomiast w Etapie II – od lipca 2018 r. do grudnia 2019 r.

Pytanie 2 Uczestnika konkursu:

Czy zbiór testowy na Etapie I będzie zawierał dokumenty zgłoszeniowe, które były dostępne publicznie - w formie oryginalnej lub jako tłumaczenie - w jakichkolwiek bazach patentowych (UPRP lub w bazach zagranicznych) przed datą 28.11.2019?

Odpowiedź Zamawiającego:

Zbiór testowy będzie zawierał wyłącznie dokumenty niepubliczne.

Pytanie 3 Uczestnika konkursu:

Czy w zbiorze testowym na Etapach I i II będą występować dokumenty każdego rodzaju, tzn.: B1, B3, T3, T4, T5, T6, Y1, S2, B2, B4 (wg "Legendy do danych")? Jeśli nie, to które dokładnie rodzaje spośród wszystkich możliwych rodzajów dokumentów będą występować?

Odpowiedź Zamawiającego:

W zbiorze testowym będą występować zgłoszenia które pokrywają się z prawami oznaczonymi B1 i Y1. Pozostałe zbiory dokumentów (poza wskazanym w zapytaniu

rodzajem S2 – dotyczącym wzoru przemysłowego!) są jak najbardziej pomocne przy uczeniu modelu.

Pytanie 4 Uczestnika konkursu:

Jakie będą proporcje ilości poszczególnych rodzajów dokumentów (p.3 powyżej) w zbiorze testowym? Czy każdy rodzaj będzie reprezentowany przez tę samą liczbę dokumentów, czy raczej pewne rodzaje są ważniejsze dla Zamawiającego i z tego względu (lub jakiegokolwiek innego) będą reprezentowane przez proporcjonalnie większą liczbę dokumentów w zbiorze testowym? Informacja ta - oraz wcześniejsze (p.1-3) - są potrzebne Uczestnikom do zapewnienia równorzędnych warunków ewaluacji modeli w trakcie treningu (eksperymenty "offline"), jak w trakcie testu.

Odpowiedź Zamawiającego:

Zamawiający upublicznił informację dla Zgłaszających jaki jest rozkład dokumentów w poszczególnych klasach MKP. Można zatem przyjąć, że prawdopodobieństwo znalezienia się dokumentu z danej klasy jest tym wyższe im częstsze są zgłoszenia. Odnośnie różnych kategorii dokumentów to zbiór testowy będzie składał się z 80 zgłoszeń wynalazków oraz 20 zgłoszeń wzorów użytkowych.

Pytanie 5 Uczestnika konkursu:

Czy dokumenty w zbiorze testowym - tak jak dokumenty treningowe - będą posiadały zawsze warstwę tekstową?

Odpowiedź Zamawiającego:

Tak

Pytanie 6 Uczestnika konkursu:

Czy klasyfikacja przypisana przez eksperta jest w każdym przypadku pełna i jedyna możliwa, w tym sensie, że inny ekspert każdorazowo przypisałby dokładnie ten sam zestaw klas MKP do zadanego dokumentu? Jeśli nie, to jak często (w przybliżeniu) zdarzają się rozbieżności między dwoma ekspertami klasyfikującymi ten sam dokument?

Odpowiedź Zamawiającego:

Klasyfikacja realizowana jest w grupie ekspertów na zasadach konsensusu (uzgadnianie stanowisk). Nie mamy zatem do czynienia z różnymi wynikowymi klasyfikacjami.

Pytanie 7 Uczestnika konkursu:

Na Etapie I, w sytuacji, gdy ekspert przydzielił więcej niż jedną klasę MKP do danego dokumentu, którą z nich uznaje się za poprawną odpowiedź systemu: pierwszą klasę spośród przydzielonych przez eksperta, czy dowolną z nich?

Odpowiedź Zamawiającego:

Na etapie pierwszym system podaje jedną klasę i musi być ona zgodna z którąś z klas wskazanych przez eksperta na 4-tym poziomie.

Pytanie 8 Uczestnika konkursu:

Czy na Etapie I lub II, system musi już posiadać funkcjonalność oznaczania fragmentów tekstu, które zadecydowały o przypisaniu takiej, a nie innej klasyfikacji? Jeśli tak, to w jaki sposób ta funkcjonalność będzie oceniana?

Odpowiedź Zamawiającego:

Nie, to nie jest przedmiotem I ani II etapu konkursu.

Pytanie 9 Uczestnika konkursu:

Jaki będzie rozmiar zbioru testowego (liczba dokumentów), na Etapie I i na Etapie II?

Odpowiedź Zamawiającego:

Około 100 dokumentów.

ZASTĘPUJĄCY DYREKTORA GENERALNEGO
Urzędu Patentowego Rzeczypospolitej Polskiej
Maria Sznuk-Olewczyńska /-/