



32-BU.610.9.2018

Toruń, dnia 29.01.2019 r.

## WYJAŚNIENIE TREŚCI ZAPYTANIA OFERTOWEGO

dot. postępowania o udzielenie zamówienia publicznego pn. *Przeprowadzenie językowej korekty warstwy tekstowej plików po maszynowym OCR.*

Uprzejmie informuję, że w dniach 23, 25 i 28 stycznia 2019 r. wpłynęły do Zamawiającego zapytania Wykonawcy dotyczące treści zapytania ofertowego w brzmieniu:

### Pytanie 1 z dnia 23.01.2019 r.:

Prosimy o podanie jednostek, według których Wykonawca ma dokonać wyceny na realizację zamówienia.

### Odpowiedź Zamawiającego:

Jednostką rozliczeniową jest jedna strona książki/czasopisma, czyli jeden skan. Przy dużej liczbie skanów oczekujemy ceny uśrednionej, w tym celu umieściliśmy próbki naszych skanów do pobrania.

### Pytanie 2 z dnia 25.01.2019 r.:

1. Prosimy o podanie procentowego podziału pomiędzy skanami czasopism a skanami książek.
2. Prosimy o podanie procentowego podziału zdigitalizowanych obiektów pomiędzy język polski, niemiecki, francuski, angielski, rosyjski, łacina oraz pozycje drukowane szwabachą.
3. W Opisie przedmiotu zamówienia podali Państwo zakresowo liczbę skanów w ilości minimum 500 tys. – maksimum 630 tys. Skąd taka duża rozbieżność w ilości skanów? Jaką ilość skanów należy przyjąć do oszacowania oferty realizacji zamówienia za maksymalną cenę brutto?
4. Uprzejmie prosimy o wskazanie tzw. wartości szacunkowej zamówienia, celem prawidłowego skalkulowania ceny i należytego przygotowania oferty.

### Odpowiedź Zamawiającego:

Przybliżone obliczenia dają:

32% skanów to czasopisma

68% skanów to książki

Przybliżone obliczenia podziału na język dają:

Obiekty z podziałem na język:

57% - polski

2% - angielski

6% - francuski

34% - niemiecki (z czego ok. połowa to szwabacha)

1% - łacina

1% - rosyjski

Rozbieżność w podanej w ofercie liczbie skanów oraz przybliżone wartości podane powyżej wynikają z tego, że proces selekcji obiektów przeznaczonych do digitalizacji trwa i trwać będzie jeszcze długo.

Maksymalna wartość jaką należy przyjąć to 630 tys. skanów; będziemy wdzięczni również za oszacowanie dolnej wartości oraz, jeśli to możliwe, kwoty za poszczególne języki.

### **Pytanie 3 z dnia 28.01.2019 r.:**

Proszę o podanie szacunkowych ilości dla poszczególnych formatów przeznaczonych do digitalizacji z podziałem na wyszczególnione w opisie języki. Zwłaszcza o ilość materiałów opisanych szwabachą.

W umowie w paragrafie 5 dotyczącym kar umownych w podpunkcie 2 b) jest: *0,5% maksymalnego wynagrodzenia brutto, o którym mowa w § 4 ust. 1 niniejszej umowy, za każdy stwierdzony przypadek nienależytego wykonania prac.* Co Zamawiający rozumie przez przypadek nienależytego wykonania prac, czy jest to stwierdzenie pojedynczego błędu w którejkolwiek z 500-630 tysięcy przeznaczonych do konwersji stronie czy niepoprawienie wykazanego błędu w wyznaczonym terminie?

### **Odpowiedź Zamawiającego:**

Na podstawie rozdziału IX Zapytania ofertowego oraz z uwagi na wydarzenia losowe (ewakuacja doraźna z budynków Zamawiającego w dniu 28.01.2019 r.) pytanie pozostało bez odpowiedzi.

### **Pytanie 4 z dnia 28.01.2019 r.:**

1. Prosimy o doprecyzowanie wymagania „Wiązki z FineReadera lub innego programu, w którym wykonano OCR”. Jakie dane ma zawierać wiązka?
2. Prosimy o informację o jakich logach jest mowa w zapisie dotyczącym formy wynikowej: „Pliki-logi z programu, w którym dokonano OCR”?
3. Prosimy o doprecyzowanie jaką nazwę powinny mieć pliki wynikowe? Czy mają mieć taką samą nazwę jak pliki wejściowe i powinny znajdować się w strukturze katalogów identycznej jak w przypadku struktury wejściowej?
4. Prosimy o doprecyzowanie zapisu: „Warstwa tekstowa musi odzwierciedlać podział tekstu w pliku, w szczególności podział tekstu na kolumny”. Czy zapis ten oznacza, że w pliku txt treść powinna być zgodna z kolejnością przedstawioną na dokumencie, a w przypadku podziału na kolumny należy zachować kolejność tekstu poczynając np. pierwszej kolumny od lewej strony?
5. Prosimy o określenie z jakich lat XIX i XX w. są powstałe książki i czasopisma? Czy myśląc o wieku XX będą znajdować się też zbiory np. z 90-tych? Z jakiego okresu lat jest najwięcej zbiorów podlegających procesowaniu?
6. W Opisie przedmiotu zamówienia oczekują Państwo „...trafność/skuteczność: 100%”. Oznacza to, że OCR w 100% jest zgodny z tekstem widocznym na cyfrowej postaci obiektu.  
Prosimy w tym miejscu o wyjaśnienie, czy jako prawidłowe wykonanie zadania uznają Państwo poprawę maszynowego OCR w zakresie znaków „niepewnych” lub „nieodczytanych przez OCR” i poprawę ich na prawidłowe? Czy weryfikacji należy poddać całość odczytanego maszynowo tekstu przez OCR łącznie ze znakami zakwalifikowanymi przez program jako poprawne?
7. W związku z faktem, iż temat jest mocno złożony a przygotowanie rzetelnej oferty wymaga czasu i przeprowadzenia odpowiednich analiz, zwracamy się z prośbą o przedłużenie terminu składania ofert do poniedziałku 4 lutego 2019 r.

### **Odpowiedź Zamawiającego:**

Na podstawie rozdziału IX Zapytania ofertowego oraz z uwagi na wydarzenia losowe (ewakuacja doraźna z budynków Zamawiającego w dniu 28.01.2019 r.) pytanie pozostało bez odpowiedzi.

Sekretarz naukowy  
Biblioteki Uniwersyteckiej  
w Toruniu  
*Liliana Lewandowska*  
dr Liliana Lewandowska