

OBLICZENIA LUDZKIE JAKO SPOSÓB NA ROZSZERZANIE BAZ WIEDZY NA PODSTAWIE GRY 'ZGADNIJ AUTORA'

Tomasz Smykowski

Uniwersytet Kazimierza Wielkiego
Instytut Mechaniki Środowiska i Informatyki Stosowanej
III rok Informatyki
ul. Chodkiewicza 30, 85-064 Bydgoszcz
email: tomasz_smykowski@o2.pl

Streszczenie: *Praca ta przedstawia wyniki gry online 'Zgadnij autora' jako przykładu gry opierającej się na zasadach human - based computation. Gra ta polega na odgadywaniu autora książki. Dzięki prostym, jasnym zasadom i dodatkowym elementom wpływającym na atrakcyjność dla graczy (nagrody, rywalizacja itp.) udało się osiągnąć znaczące efekty w gromadzeniu bazy danych połączeń autor – książka. Są to dane dostępne tradycyjnymi sposobami, ale ciężko osiągalne. Oprócz tego udało się wypracować mechanizmy i sprecyzować założenia jakie gry human - based computation muszą spełniać, aby były atrakcyjne dla użytkowników oraz jak zapobiegać dodawaniu błędnych danych do bazy.*

Słowa kluczowe: : human - based computation, obliczenia ludzkie, gra, zgadnij autora, baza wiedzy, autor, książka

Human - based computation as a way to broaden knowledge bases. The case of the 'Guess the author' game

Abstract: *Abstract: This article presents the results of an on-line game called "Guess the author" as an example of a game grounded on human - based computation. The game consists in guessing authors of books. Thanks to its easy and clear rules as well as additional features that make the game more attractive to players (like awards, competition etc.), the game has been successful in collecting a data base of author – book links. This data is available in conventional way but rather hard to achieve. Apart from that, the authors successfully developed mechanisms and specified conditions that have to be met by human - based computation games so as to make them attractive and to prevent adding incorrect data to the database.*

Keywords: human - based computation, game guess the author, knowledge base, author, book

1. WPROWADZENIE

Technika *human – based computation* polega na przekazywaniu ludziom części obliczeń do wykonania przez komputery, które same nie mogą ich zrealizować. Dzięki połączeniu celów naukowych i atrakcyjności narzędzi można połączyć dużą liczbę osób w celu powiększania bazy danych o informacje, które w żaden inny sposób nie mogą zostać zgromadzone. Dane takie nie mogą zostać wprowadzone przez narzędzia automatyzacji lub przez inne algorytmy komputerowe w sposób efektywny. [1]

Gra 'Zgadnij autora' została umieszczona na stronie internetowej założonej przez autora. Została ona napisana w języku *PHP* z użyciem bazy danych *MySQL*.

Gra polega na odgadywaniu autorów książek podawanych przez innych użytkowników.

Celem podstawowym gry jest analiza możliwości human – based computation w zastosowaniu w następnych przedsięwzięciach związanych z gromadzeniem wiedzy, m.in. na temat hipernimii i znaczenia wyrazów. Powiązania autor – książka są możliwe do uzyskania przez narzędzia automatyzacji, ale jest to utrudnione. Dlatego autor postanowił użyć takiego celu do sprawdzenia możliwości metody human – based computation.

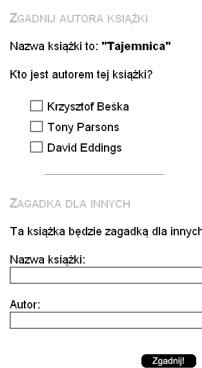
2. ZASADY GRY

Ilustracja 1 przedstawia formularz gry. Składa się on z 4 części, przy czym część 'Zgadnij autora książki' i 'Zgadka dla innych' dotyczą samej rozgrywki. Po lewej

stronie w części 'Wynik' znajdują się informacje statystyczne oraz 'Zasady gry' ze zbiorem reguł gry dla początkujących graczy. Jak widać na załączonych ilustracjach oprawa graficzna gry jest czytelna i atrakcyjna wizualnie dla odbiorcy.



Rys. 1 Widok na cały ekran gry 'Zgadnij autora'.



Rys. 2 Przykładowy formularz rozgrywki, który mogą zobaczyć uczestnicy gry.

Zasady gry są następujące:

Autorów trzeba podawać bez skracania imion. Jeżeli nie ma polskiego odpowiednika, trzeba podać nazwę angielską. Za każdego odgadniętego autora otrzymuje się 1 punkt. Za nie odgadniętego autora nie traci się punktów. Jeżeli w zagadce jest błąd, trzeba nacisnąć "Zgadnij", aby przejść do następnego pytania. W przypadku nieznamości odpowiedzi należy odświeżyć stronę, aby przejść do kolejnego pytania. Brak odpowiedzi na pytanie zagadki nie powoduje utraty punktów. Grę można kontynuować w każdej chwili, nawet po ponownym zalogowaniu się na stronę. W celu zapewnienia jakości gry, konta osób naruszających ciężko powyższe zasady są blokowane.

Zasady te zostały wypracowane w trakcie udziału w grze uczestników. Są one kompromisem między wymaganiami systemu gromadzenia danych, możliwościami ich automatycznej obróbki i atrakcyjności gry dla uczestników, co zostanie objaśnione szczegółowo w rozdziale 4. Warto tutaj zauważyć, że przejrzystość zasad jest istotna aby uniknąć kontrowersji wobec ich sprawiedliwości. [2], [3]

Użytkownik ma dostęp do statystyk dostarczających informacji o tym, na którym miejscu znajduje się w rankingu gry oraz ile uzyskał punktów. Na stronie rankingu znajduje się informacja o 10 najlepszych użytkownikach od początku trwania gry i o 10 użytkownikach najlepszych bieżącego dnia:

NAJLEPSI DZISIAJ		
Miejsce	Użytkownik	Poprawnych odpowiedzi
1	pearl	5

NAJLEPSI Z NAJLEPSZYCH		
Miejsce	Użytkownik	Poprawnych odpowiedzi
1	marcysia	1055
2	bejs2	892
3	Effa	549
4	kiniunia25	504
5	lakie	405
6	peka	274
7	Ela	270
8	ghwan	248
9	ala	245
10	renataszagabka	176

Rys. 3 Ranking gry.

3. SZCZEGÓLOWE CELE GRY

Cele gry dzielą się na dwa rodzaje. Pierwszy z nich dotyczy tego, o jakie dane ma zostać wzbogacona baza danych przez udział uczestników. Drugi rodzaj to cele, które gra realizuje dla samych uczestników i wpływa na ich satysfakcję z udziału w niej. Celem gry z punktu widzenia autora jest zgromadzenie bazy książek, czyli połączeń autor – książka. Największa strona o książkach w Polsce posiada przeszło 30 tys. autorów i 90 tys. książek. Autor postawił za cel do realizacji otrzymanie 10 tys. wpisów samoweryfikujących się. Zakładając, że 80% wpisów będzie weryfikować 20% pozostałych. Założono zatem osiągnięcie 2 tys. poprawnych i zweryfikowanych powiązań autor – książka, co spełniłoby *zasadę Pareto*. [4]

Celem gry dla użytkowników jest chęć weryfikacji swojej wiedzy na temat książek. Sprawdzenie, jak dużo książek się pamięta. Cel ten jest realizowany przez udział w grze i otrzymane statystyki. Celem jest realizacja i porównanie swojej wiedzy z innymi uczestnikami gry – do osiągnięcia tego służy ranking 10 najlepszych uczestników. W trakcie gry dodano także ranking 10

najlepszych uczestników bieżącego dnia w celu pozytywnego wzmocnienia motywacji uczestników, którzy na początku muszą zdobyć relatywnie dużo punktów, aby trafić do głównego rankingu.

Celem gry jest czysta zabawa – udział w grze nie wymaga żadnych specjalnych umiejętności poza wiedzą o powiązaniach książka – autor. Oprócz tego w trakcie gry podjęto próby znalezienia sponsorów nagród w celu zwiększenia atrakcyjności gry dla użytkowników. Uzyskano wsparcie w postaci nagród rzeczowych.

Aby uzyskać odpowiednio dużą bazę samo weryfikujących się danych gra musi spełniać jak najlepiej cele stawiane przez użytkowników [5]. W ten sposób gra jest dla nich atrakcyjna i chętnie biorą w niej udział. W trakcie tworzenia gry napotkano na problem związany ze zwiększeniem popularności pomysłu i samej gry, co wymusiło usprawnienie szybkości działania procesów weryfikacji danych. Jednak liczba 239 zarejestrowanych użytkowników w kontekście tych problemów jest odczytywana jako sukces. Pod uwagę trzeba wziąć też, na czym polega gra – mimo skierowania jej do wąskiej grupy osób, osiągnięto dosyć dużą bazę danych. Baza ta zostanie opisana w dalszej części tego opracowania.

4. PROBLEMY I ROZWIĄZANIA

W czasie gry rozpoznano kilka problemów związanych z wprowadzaniem danych przez użytkowników. Pierwotnie gra polegała na podawaniu autora w polu tekstowym. Z tego powodu, aby dostać punkt trzeba było podać dokładnie takie same nazwisko jak autor zagadki. Z powodu błędów dokonywanych przez użytkowników ten sposób nie okazał się skuteczny. Zostało to zastąpione wyborem z listy trzech autorów z powodu możliwych błędów w pisowni autorów.

Oprócz tego pojawiły się problemy ze skracaniem przez użytkowników nazw autorów albo nieznanomość ich pełnych nazwisk. Problem ten został rozwiązany przez rozszerzenie regulaminu i dodanie blokady dodawania autorów ze skróconymi nazwiskami.

Następny problem to wprowadzanie danych w złe pola: w pole autora – nazwy książki, a w pole nazwy książki – autora. Pojawiły się także niedozwolone znaki – np. nawiasy klamrowe, cudzysłowie itp. Dodatkowo pojawiały się wpisy bez polskich liter, drobne literówki, nieużywanie dużych liter w nazwach książek i nazwiskach autorów. Te problemy zostaną rozwiązane na etapie automatycznej weryfikacji.

Oprócz powyższych pojawili się użytkownicy, którzy dodający fałszywe dane, tylko w celu zanieczyszczenia bazy danych i uzyskania wysokich miejsc w rankingu. Aby zabiec temu zjawisku przewidziano następujące

środki zaradcze: aby wziąć udział w grze trzeba się zarejestrować, natomiast aby dodać nową książkę, trzeba odgadnąć autora innej książki.

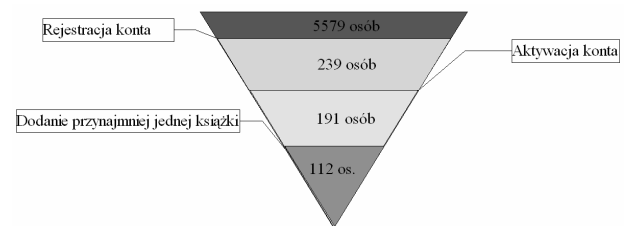
Drugi mechanizm jest związany też z samo weryfikacją danych, naturalną konsekwencją tego, że gra sama w sobie nie jest łatwa, jest to, że użytkownicy wprowadzający fałszywe dane szybko sami rezygnują z tych działań. Przygotowane narzędzia automatyczne i pół-automatyczne powodują wychwytywanie takich użytkowników i blokowanie im kont, albo usuwanie błędnych wpisów, jeżeli są one sporadyczne.

Samo weryfikacja wpisów polega na tym, że można na podstawie zebranych danych automatycznie wybrać poprawne wpisy. Elementami samo weryfikacji są:

- ile osób niezależnie wprowadziło takiego samego autora i nazwę książki
- ile osób odgadło / nie odgadło poprawnie autora podanej książki
- ile poprawnych książek wprowadził użytkownik do tej pory

5. WYNIKI PRZEDSIĘWZIĘCIA

Przedsięwzięcie trwało od 25 grudnia 2007 i trwa nadal. Powyższe opracowanie jest przygotowane na podstawie danych z dnia 1 maja 2008, czyli z 5 miesięcznego okresu czasu. Stronę odwiedziło 5579 internautów, zarejestrowało się 239 użytkowników. Spośród tych osób 48 nie aktywowało swojego konta przez kliknięcie w link aktywacyjny w wiadomości email, którą otrzymali.



Rys. 4 Odwrócona piramida liczby osób, które podjęły działania związane z grą.

Z sumy 239 użytkowników z aktywowanymi kontami w grze uczestniczyło $239 - 127 = 112$ osób. Oznacza to, że tylko 46% osób, które założyły konto na stronie aktywowało swoje konto i dodało przynajmniej jedną książkę. Średnia ta sprawdzała się przez większość czasu trwania gry.

Co druga osoba zrezygnowała z udziału, Wykluczone zostały więc osoby, które nie są przekonane i zdecydowane brać udziału w grze, której zasady zostały ściśle określone. Z punktu widzenia celów naukowych