

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Krzysztof Dudzik

Nr albumu: 248349

**Aplikacja wspomagająca
tworzenie i edycję haseł
w polskim Wikisłowniku**

**Praca magisterska
na kierunku INFORMATYKA**

Praca wykonana pod kierunkiem
dr. hab. Jerzego Tyszkiewicza, prof. UW
Instytut Informatyki

2011

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Oświadczenie autora pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora pracy

Streszczenie

Tematem pracy jest aplikacja służąca do ułatwienia pracy autorów haseł w polskim Wikisłowniku. Jej funkcje mają w maksymalny możliwy sposób ułatwić tworzenie i edytowanie haseł osobom bez wiedzy informatycznej i technicznej, a także automatyzować możliwie wiele rutynowych czynności wykonywanych przy redagowaniu hasła, jak tworzenie łączy do haseł powiązanych, zautomatyzowane szukanie przykładów użycia, wystąpień w związkach frazeologicznych, wyrazów bliskoznacznych, innych słów, którą formę gramatyczną mogłoby stanowić hasło itp. Dodatkowo aplikacja może przejąć część funkcji realizowanych obecnie za pomocą botów.

Słowa kluczowe

Wikisłownik, Fundacja Wikimedia, MediaWiki, wiki, edytor, API, JavaScript, jQuery, interfejs użytkownika, społeczność internetowa, open source

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.3 Informatyka

Klasyfikacja tematyczna

H. Information Systems

H.5. Information Interfaces and Presentation

H.5.3. Group and Organization Interfaces

Tytuł pracy w języku angielskim

An application supporting article creation and edition for the Polish Wiktionary

Spis treści

1 Wprowadzenie	5
2 Wikisłownik	7
2.1 Projekty Fundacji Wikimedia	7
2.2 Oprogramowanie MediaWiki	9
2.2.1 Edytowanie i wikitekst	10
2.3 Wiktory – Wikisłownik	12
2.4 Polska edycja Wikisłownika	14
2.4.1 Struktura hasła	15
2.4.2 Wady dotychczasowych rozwiązań	26
2.4.3 Różnice w stosunku do Wikipedii	29
3 Aspekty społecznościowe	31
3.1 Koncepcja <i>wiki</i>	31
3.2 Społeczność polskiej edycji Wikisłownika	32
3.3 Analiza wymagań	33
3.3.1 Wymagania funkcjonalne	33
3.3.2 Wymagania нефункционалне	33
3.3.3 Spodziewane korzyści	33
3.4 Specyfika tworzenia aplikacji dla wikispołeczności	33
4 Opis implementacji	35
4.1 Wprowadzenie	35
4.2 Formularz edycyjny	37
4.3 Automatyzacja edycji hasła	37
4.4 Wdrożenie i dalszy rozwój	37
5 Podsumowanie	39

Bibliografia	41
Spis ilustracji	44
Spis tabel	44

Rozdział 1

Wprowadzenie

Żyjemy w czasach, w których nieustannie zmienia się sposób wyszukiwania informacji przez przeciętnego człowieka. Z roku na rok coraz mniejszą rolę odgrywają papierowe kompendia takie jak encyklopedie i słowniki, stopniowo przybierają natomiast na znaczeniu elektroniczne bazy wiedzy – szczególnie zaś internetowe zbiory danych. Przyczyny tego stanu rzeczy są oczywiste: chodzi przede wszystkim o wygodę korzystania ze stron internetowych. Brak możliwości wyszukiwania w obrębie ogromnych ilości danych powoduje, że encyklopedie i słowniki w postaci książek stają się o wiele mniej atrakcyjne dla kogoś, kto chce zdobyć nowe informacje.

Wszechobecny dostęp do internetu sprawia, że to właśnie w sieci WWW powstają najbardziej popularne bazy ludzkiej wiedzy. Nie ma chyba internautów, którzy nie korzystali by, rzadziej lub częściej, z Wikipedii – internetowej encyklopedii pisanej przez ochotników. Właśnie fakt, że encyklopedia ta współtworzona jest przez amatorów, stanowi o jej wyjątkowym charakterze, który zostanie w tej pracy pokrótce opisany. Wikipedia stale utrzymuje się w pierwszej dziesiątce najczęściej odwiedzanych stron, a pod wieloma względami jest to dziś najlepsza istniejąca encyklopedia. Przed kilkoma laty głośne było porównanie jej z prestiżową *Encyclopædia Britannica* – okazało się, że różnice w poziomie merytorycznym są niewielkie.

O ile przewrót w kategorii encyklopedii właściwie już się dokonał, nieco inaczej wygląda rywalizacja słowników. Oczywiście wyraźnie widać, że i tu papierowe edycje są coraz mniej popularne. Różnice uwidaczniają się, gdy przeanalizowana zostanie sytuacja słowników internetowych. Tak zwany siostrzany projekt Wikipedii, Wikisłownik, nie dominuje wśród

konkurencji – zarówno na świecie, jak i w Polsce. Przyczyny tego stanu rzeczy są złożone. Autor postanowił skupić się na kilku zagadnieniach, uwidaczniających się w polskojęzycznej wersji Wikisłownika. W tym celu konieczne było zbadanie społeczności zaangażowanej w tworzenie tego projektu. Jego efektem było wykonanie prac programistycznych, których opis stanowi główną część niniejszego opracowania.

W przypadku wszystkich projektów opartych na silniku programistycznym MediaWiki istotną barierą rozwoju jest sama technologia. Każdy ochotnik ma możliwość uczestniczenia w rozwoju portalu, wiąże się to jednak z koniecznością przystosowania się do wymagań stawianych przez oprogramowanie. Edytowanie haseł w internetowej encyklopedii czy słowniku jest praktycznie niemożliwe dla osoby bez wcześniejszego przygotowania lub znacznej wiedzy techniczno-informatycznej. Oprogramowanie MediaWiki oparte jest bowiem na tzw. wikikodzie (także: wikitekst, wikiskładnia), czyli języku opisu struktury i wyglądu strony internetowej – prostszym niż HTML, jednak wciąż nieintuicyjnym dla kogoś, kto nie miał wcześniej do czynienia z tego typu edytorami. Dlatego wielu potencjalnych współautorów zniechęca się do projektu już przy pierwszej próbie poprawy artykułu.

Aby zmienić tę sytuację, przygotowany został nowy edytor, dostosowany specjalnie do potrzeb polskiego Wikisłownika. Aplikacja pozwala na o wiele prostsze tworzenie nowych i zmienianie starych haseł niż poprzednia, standardowa. Dzięki użyciu jej jako domyślnej w projekcie popularyzacja edytowania Wikisłownika wśród fachowców w dziedzinach lingwistycznych okaże się łatwiejsze – zniknie podstawowa bariera, jaką jest konieczność dostosowania się do skomplikowanych technicznych wymagań stawianych przez użyte oprogramowanie. Dodatkowo nowa aplikacja umożliwia zaawansowaną automatyzację tworzenia hasła. Wiele z czynności zintegrowanych z nowym edytorem do tej pory wymagało mozolnych poszukiwań w artykułach Wikisłownika oraz innych projektach. Dzięki użyciu API udostępnianego przez serwisy Fundacji Wikimedia skomplikowane przeszukiwanie tysięcy stron udało się sprowadzić do kilku kliknięć.

W dalszej części pracy opisany został proces tworzenia tego edytora. Pierwszy rozdział charakteryzuje pokrótce sam Wikisłownik, jak i pokrewne projekty oraz oprogramowanie w nich użyte. Następnie opisano społecznościowe aspekty tworzenia tego typu aplikacji ze szczególnym uwzględnieniem koncepcji *wiki*. Ostatni rozdział wyczerpująco przedstawia szczegóły projektowe i implementacyjne aplikacji.

Rozdział 2

Wikisłownik

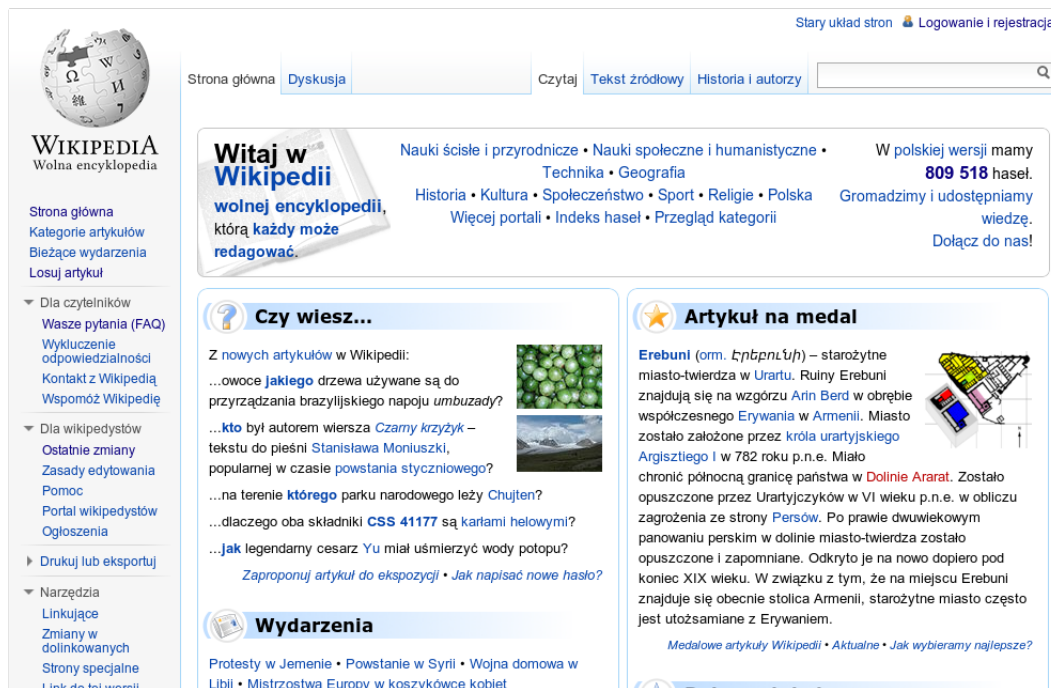
Rozdział ten stanowi charakterystykę Wikisłownika – sieciowego słownika opartego na oprogramowaniu MediaWiki. Wikisłownik jest jednym z największych i najpopularniejszych słowników dostępnych w polskim internecie. W kolejnych sekcjach projekt ten został opisany na różnych poziomach szczegółowości. Omówiono zarówno oprogramowanie, na jakim bazuje słownik, jak i swego rodzaju „ekosystem”, w którym znajduje on swoje miejsce.

2.1. Projekty Fundacji Wikimedia

Podmiotem odpowiedzialnym m.in. za rozwój Wikisłownika jest Wikimedia Foundation Inc. (opisywana dalej jako „Fundacja”) – organizacja non-profit mająca siedzibę w San Francisco w Stanach Zjednoczonych, istniejąca od 2003 roku. Jak informuje strona internetowa polskiego partnera Fundacji, Stowarzyszenia Wikimedia Polska, *celem fundacji jest sprzyjanie tworzeniu i rozwojowi projektów o otwartej treści opartych na technologii WikiWiki oraz dostarczanie społeczności internetowej pełnej zawartości wymienionych projektów za darmo i bez zamieszczania reklam* [1]. Doskonale znaną, sztandarową inicjatywą Fundacji jest Wikipedia (<http://www.wikipedia.org>) – największa obecnie encyklopedia internetowa, dostępna w 281 językach (stan z sierpnia 2011 roku) i zawierająca ponad 19 milionów haseł, w tym ponad 3,7 miliona w największej, angielskojęzycznej¹ edycji [2]. Mimo częstej krytyki tego

¹ Oficjalnie w projektach Fundacji używane są określenia typu *angielskojęzyczny*, *polskojęzyczny*. Choć w przypadku wersji polskojęzycznej znakomita większość uczestników projektów pochodzi z Polski, nie jest to regułą dla innych edycji. W dalszej części pracy przyjęto uproszczenie polegające na tym, że określenia typu *polska*

przedsięwzięcia faktem jest, że Wikipedia jest miejscem, z którego miliony osób korzystają, by pozyskać informacje z najróżniejszych dziedzin. Obecny stan rzeczy możliwy jest dzięki pracy wielkiej liczby wolontariuszy tworzących artykuły bez wynagrodzenia.



Ilustracja 2.1: Polska edycja Wikipedii

Wikipedia jest najbardziej znanym, ale nie jedynym projektem pod opieką Fundacji. Pozostałe to tzw. „projekty siostrzane”, w szczególny sposób uwzględniane również przy tworzeniu haseł w encyklopedii. Oto lista wspieranych przez Fundację wielojęzycznych inicjatyw [3]:

- Wikisłownik (ang. *Wiktionary*) – wielojęzyczny słownik internetowy, będący głównym przedmiotem niniejszej pracy,
- Wikicytaty (ang. *Wikiquote*) – zbiór cytatów autorstwa znanych osób, z filmów i książek, przysłów i porzekadeł,
- Wikibooks – serwis z „otwartymi” (opartymi na wolnej licencji) podręcznikami,
- Wikiźródła (ang. *Wikisource*) – zbiór dokumentów źródłowych w wersjach oryginalnych i tłumaczonych, nieograniczonych prawem autorskim,
- Wikinews – otwarty serwis informacyjny,

Wikipedia, angielski *Wikisłownik* traktowane są jako tożsame z określeniami używanymi sformułowań z częstką -języczny.

- Wikiversity – materiały edukacyjne i naukowe,
- Wikispecies – katalog gatunków organizmów żywych,
- MediaWiki – nadzór nad tworzeniem i rozpowszechnianiem oprogramowania MediaWiki (p. sekcja 2.2),
- Meta-Wiki – projekt ułatwiający koordynację wszystkich pozostałych.
- Wikimedia Commons – repozytorium mediów (zdjęć, grafik, filmów) dostępnych na wolnej licencji, z którego korzystają pozostałe projekty Wikimedia,
- Wikimedia Incubator – metaprojekt umożliwiający tworzenie nowych inicjatyw wspieranych przez Fundację.

Wszystkie projekty łączy sposób ich powstawania – możliwość edycji dostępna jest praktycznie dla każdego internauty. Nie dotyczy to co prawda kilku krajów, w których projekty Fundacji zablokowane są w ramach cenzury internetu, jednak ogromna większość osób dysponujących łączem internetowym ma szansę stać się jednymi spośród współautorów haseł.

Drugą cechą wspólną są wolne licencje, na których udostępniana jest zawartość wszystkich serwisów. Po reformie w czerwcu 2009 roku treść Wikipedii i projektów siostrzanych dostępna jest nie tylko na licencji GNU FDL (Free Documentation License), ale także na kompatybilnej z nią CC-BY-SA 3.0 (Creative Commons Attribution-ShareAlike / Uznanie Autorstwa – Na Tych Samych Warunkach) [4]. Oznacza to, że można ją dowolnie wykorzystywać we własnych dziełach pod warunkiem podania oryginalnych autorów i zachowania pierwotnej licencji.

2.2. Oprogramowanie MediaWiki

Sama działalność wolontariacka redaktorów projektów Wikimedia nie wystarczyłaby do stworzenia serwisów internetowych o obecnych kształtach. Konieczne jest oczywiście również zapewnienie oprogramowania, które umożliwi płynną współpracę przy tworzeniu haseł. Tym oprogramowaniem jest wolna platforma MediaWiki tworzona zgodnie z zasadami *open source*. System MediaWiki napisany jest w języku PHP i obsługuje kilka popularnych baz danych (w przypadku projektów Wikimedia jest to MySQL). Dla inicjatyw Wikimedia

stanowi szkielet programistyczny od samego ich początku, a od 2002 roku stale się rozwija. W czerwcu 2011 roku wersją używaną w projektach było MediaWiki 1.17.

System MediaWiki używany jest nie tylko w projektach wspieranych przez Fundację, ale także w tysiącach innych, mniejszych lub większych, co jest możliwe dzięki wysokiemu stopniowi konfigurowalności i dużej liczbie dostępnych rozszerzeń. Są to w dużej mierze serwisy o podobnym charakterze, umożliwiające swobodną wymianę informacji na dowolny temat. MediaWiki bywa także używane w firmowych intranetach i wszędzie tam, gdzie zachodzi potrzeba udostępnienia materiałów do edycji dużej liczbie użytkowników.

2.2.1. Edytowanie i wikitekst

Strony w projektach opartych na platformie MediaWiki na ogół nie mogą być czystym tekstem, pozbawionym formatowania. Przykładowo hasła w encyklopedii muszą zachowywać określoną strukturę – występuje więc podział na sekcje, ilustracje, różne rodzaje formatowania (kursywa, wytłuszczenie), przypisy czy powtarzalne fragmenty. Szczególnie istotnym elementem są linki pomiędzy poszczególnymi artykułami, wyróżniające projekty Fundacji na tle ich papierowych, ale też elektronicznych konkurentów. Odnośniki pozwalają błyskawicznie przemieszczać się między hasłami, by w ten sposób uzyskiwać kolejne informacje wspomagające przyswajanie wiedzy.

Linki i formatowanie na stronach internetowych tworzone są za pomocą elementów języka HTML lub XHTML. O ile języki te są proste w obsłudze dla specjalisty informatyka, to laik nie jest w stanie stworzyć za ich pomocą stron bez uprzedniego dłuższego przygotowania. Aby umożliwić bezproblemową edycję stron internetowych osobom bez wykształcenia informatycznego, programiści MediaWiki zaprojektowali tzw. wikitekst [5] – uproszczony język opisu stron, pozwalający na realizację wymienionych elementów. Porównanie niektórych z nich znajduje się w tabeli 2.1.

Łatwo można zauważyć, że używanie wikitekstu jest o wiele prostsze niż nauka języka XHTML. Jeśli zachodzi potrzeba zaawansowanego formatowania, możliwe jest także użycie znaczników XHTML. W przypadku standardowego formatowania jest to jednak niewskazane ze względu na dobro niedoświadczonych edytorów.

Bardzo istotnym elementem wikitekstu są szablony – predefiniowane fragmenty kodu z opcjonalnymi parametrami. Szablony można uznać za odpowiednik procedur/funkcji

	Wikitekst	XHTML
Kursywa	<code>''Tekst''</code>	<code>Tekst</code>
Wytłuszczenie	<code>'''Tekst'''</code>	<code>Tekst</code>
Nagłówek	<code>== Nagłówek ==</code> <code>=== Nagłówek ===</code>	<code><h2>Nagłówek</h2></code> <code><h3>Nagłówek</h3></code>
Odnośnik wewnętrzny	<code>[[Strona]]</code> <code>[[Strona strony]]</code>	<code>Strona</code> <code>strony</code>
Odnośnik zewnętrzny	<code>[http://www.google.com Google]</code>	<code>Google</code>
Obraz	<code>[[Plik:Przykład.png thumb Podpis]]</code>	<code>
</code> <code><div class="caption">Podpis</div></code>
Podział na akapity	Pierwszy akapit Drugi akapit	<code><p>Pierwszy akapit</p></code> <code><p>Drugi akapit</p></code>
Lista nienumerowana	<code>* Element</code> <code>* Element</code> <code>* Element</code>	<code></code> <code>Element</code> <code>Element</code> <code>Element</code> <code></code>
Lista numerowana	<code># Element</code> <code># Element</code> <code># Element</code>	<code></code> <code>Element</code> <code>Element</code> <code>Element</code> <code></code>

Tabela 2.1: Porównanie HTML i wikitekstu

w językach programowania. W projektach opartych na MediaWiki szablony pełnią przede wszystkim dwie główne funkcje:

- upraszczają kod – pozwalają np. na zastąpienie skomplikowanego kodu XHTML (a także jeszcze bardziej złożonych funkcji parsera MediaWiki) krótkim wywołaniem szablonu,
- standaryzują strony – często wykorzystywane fragmenty wywoływane są zawsze w dokładnie ten sam sposób.

We wszystkich większych projektach Fundacji szablony są bardzo często wykorzystywane. Przy tym stanowią duże ułatwienie dla technicznie zaawansowanych autorów, którzy wspomagają się przy tworzeniu haseł dodatkowymi technologiami. Na używaniu szablonów korzystają przede wszystkim boty, czyli programy dokonujące edycji samodzielnie po uprzednim przygotowaniu lub pod stałą opieką programisty. W większości projektów przyjęte jest, że każdy bot ma własne konto użytkownika, nie używa natomiast konta swojego „właściciela”. Dzięki botom możliwe jest np. masowe tworzenie haseł w Wikipedii na

ściśle określony temat, jeśli istnieje dobre źródło w formie czytelnej dla komputera (takich jak opisy asteroid czy wszystkich miejscowości lub jednostek administracyjnych w danym kraju). Innym ich zastosowaniem jest automatycznie uzupełnianie tzw. interwiki – czyli odnośników pomiędzy poszczególnymi wersjami językowymi tego samego hasła.

Szablony mogą być wykorzystywane w prosty sposób nawet przez początkujących użytkowników – aby wywołać szablon, wystarczy wpisać jego nazwę pomiędzy podwójnymi nawiasami klamrowymi ({{Nazwa szablonu}}) lub {{Nazwa szablonu|parametr=abc}}). W polskim Wikisłowniku szablony pełnią szczególną rolę – i to m.in. dzięki ich szerokiemu zastosowaniu w projekcie zrodził się pomysł na niniejszą pracę. Szczegóły tego zagadnienia zostaną przedstawione w sekcji 2.4.

2.3. Wiktionary – Wikisłownik

The screenshot shows the Polish edition of Wiktionary. At the top left is a logo with various characters. Below it is the text 'Wolny, wielojęzyczny WIKISŁOWNIK'. To the right of the logo is a navigation menu with links like 'Strona główna', 'Indeks alfabetyczny', 'Spis języków', etc. The main content area has a header with 'Witaj w Wikisłowniku, wielojęzycznym słowniku!' and a welcome message. Below this are three main sections: 'O Wikisłowniku' (About Wiktionary), 'Indeksy Wikisłownika' (Wiktionary Indexes), and 'Słowo tygodnia' (Word of the Week). The 'Indeksy Wikisłownika' section shows statistics like 'ponad 248 886 haseł' and 'na 223 617 stronach'. The 'Słowo tygodnia' section features the word 'hiacynt' (hyacinth) with its definition and a small image of the flower.

Ilustracja 2.2: Polska edycja Wikisłownika

Jednym z największych projektów siostrzanych Wikipedii jest Wikisłownik, w wersji angielskiej (i wielu innych) noszący nazwę *Wiktionary* (<http://www.wiktionary.org>). Ten słownik internetowy nie rozwinął się jeszcze tak pręźnie jak encyklopedia, zwłaszcza jeśli chodzi o polską edycję. Jest dziś jednak jednym z największych słowników w sieci, a w pewnych

zastosowaniach stanowi najlepszy wybór. Dużą zaletą Wikisłownika jest jego wielojęzyczność – w tym samym serwisie znaleźć można hasła w ponad 250 językach. W przypadku niektórych z nich jest to praktycznie jedyny słownik internetowy lub nawet jedyny dostępny słownik w ogóle. Przykładem może być polski Wikisłownik, który zawiera prawdopodobnie jedyny polski słownik języka hawajskiego czy największe słowniki języków suahili i jidysz [6].

Wspomniane zostało zastosowanie szablonów do standaryzacji kodu źródłowego i struktury haseł. Trzeba jednak zaznaczyć, że dotyczy to wyłącznie haseł w obrębie jednej wersji językowej Wikisłownika, są one bowiem niezależne od siebie i od Fundacji. Wspólne dla wszystkich edycji jest jedynie oprogramowanie MediaWiki i umiejscowienie na serwerach Fundacji pod adresem `xxx.wiktionary.org`, gdzie zamiast `xxx` wstawiany jest dwu- lub trzyliterowy skrót ISO języka (np. `pl.wiktionary.org` = język polski, `de.wiktionary.org` = język niemiecki, `sq.wiktionary.org` = język albański, `csb.wiktionary.org` = język kaszubski). Wszystkie kwestie organizacyjne w obrębie wersji językowej ustalane są w ramach dyskusji i głosowań przez internetową społeczność. Głosowania służą także wyborowi administratorów projektu, czyli użytkowników mających dodatkowe uprawnienia, spośród których najważniejsze to usuwanie i zabezpieczanie haseł oraz blokowanie użytkowników działających na szkodę projektu. W lipcu 2011 roku w angielskim Wikisłowniku działało aktywnie 76 administratorów [7], zaś w polskim – 14 [8]. Dla porównania Wikipedia w języku angielskim ma 1541 administratorów (niekoniecznie aktywnych), wersja polska natomiast 163 [9].

Podobnie jak w przypadku Wikipedii, hasła w poszczególnych wersjach językowych Wikisłownika są łączone poprzez odnośniki interwiki, znajdujące się na dole lewego menu w większości haseł. W stosunku do encyklopedii występuje znacząca różnica w sposobie funkcjonowania tych linków. Wikipedia poprzez interwiki łączy artykuły na ten sam temat, często różniące się tytułem (polski artykuł *Kot domowy* odsyła do angielskiego *Cat* czy niemieckiego *Hauskatze*). W Wikisłowniku mechanizm interwiki łączy zaś strony o tym samym tytule, niezależnie od ich zawartości. Polskie hasło *kot* zawiera więc łącza do haseł o tytule *kot* w innych językach, które mogą, ale nie muszą zawierać m.in. objaśnienia polskiego znaczenia tego słowa. Z tego względu łatwo jest odnaleźć brakujące informacje o wybranym słowie w danym języku, jeśli skorzysta się z kilku edycji językowych Wikisłownika. Natomiast tłumaczenia tytułu hasła na inne języki znajdują się bezpośrednio w treści artykułu,

w miejscu przeznaczonym dla takich informacji.

Choć poszczególne wersje Wikisłownika różnią się między sobą, wspólna jest struktura hasła na najogólniejszym poziomie. Każde hasło jest podzielone na sekcje nagłówkami 2. stopnia (p. tabela 2.1), podobnie jak w innych projektach Fundacji. W przypadku słownika sposób podziału artykułu został jasno sprecyzowany i jest wspólny dla wszystkich edycji. W każdej z sekcji objaśniono znaczenie tytułowego hasła w innym języku. Przykładowo hasło *nie* w większości wersji językowych ma sekcje objaśniające znaczenie m.in. w języku polskim oraz języku niemieckim (*nigdy*). Właśnie ta ogólna cecha struktury haseł pozwala na częściową automatyzację niektórych często wykonywanych w trakcie edycji czynności.

Aplikacja opisana w niniejszej pracy przeznaczona jest dla polskiej edycji Wikisłownika. Konieczne jest zatem scharakteryzowanie specyfiki tego projektu (p. sekcja 2.4). Ze względu na odmiennosc poszczególnych wersji językowych prawdopodobnie w innych nie będzie można wykorzystać aplikacji. Z pewnością może ona być dla nich jednak bardzo przydatna – duża część funkcji przez nią realizowanych może być używana niezależnie od specyfiki danej wersji, a budowa aplikacji pozwoli na ich wyodrębnienie.

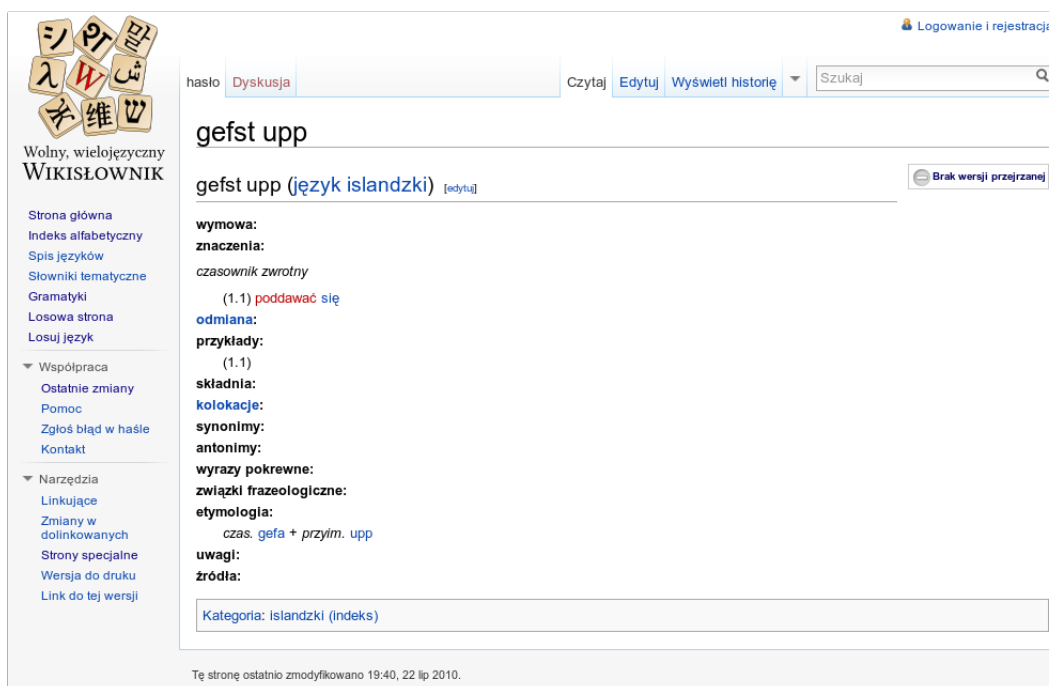
2.4. Polska edycja Wikisłownika

Wikisłownik w języku polskim powstał 22 marca 2004 r. [10] i jest dziś jedną z najlepiej rozwiniętych edycji. Jeśli chodzi o liczbę haseł, polska wersja zawiera ich ponad 230 000 i znajduje się na 8. miejscu (pierwsze trzy miejsca zajmują z dużą przewagą edycje angielska, francuska i chińska). Warto zwrócić jednak uwagę, że liczba artykułów nie musi odzwierciedlać ogólnego poziomu rozwoju danej edycji, a przynajmniej w o wiele mniejszym stopniu niż w Wikipedii. Automatyczne importowanie haseł słownikowych z innych źródeł jest prostsze niż w przypadku wpisów do encyklopedii, a hasła takie oczywiście nie wyróżniają się pozytywnie pod względem jakości. Wydaje się, że lepszym wskaźnikiem rozwoju są np. liczby aktywnych użytkowników lub administratorów. Tutaj polski Wikisłownik okazuje się wielokrotnie lepszy niż edycje litewska czy malajska, górujące nad nim pod względem liczby haseł [11].

W kolejnej podsekcji opisana została dokładnie struktura haseł w polskim Wikisłowniku. Społeczność tego projektu scharakteryzowano natomiast w sekcji 3.2.

2.4.1. Struktura hasła

Jak wspomniano, wspólny dla wszystkich wersji językowych Wikisłownika jest podział na sekcje, odpowiadające poszczególnym językom. Polska edycja Wikisłownika jest bardzo dobrze ustrukturyzowana w dalszym zakresie, co niekoniecznie jest regułą dla pozostałych edycji. Wynikiem tego jest jednolity wygląd hasła, który otrzymuje czytelnik, oraz jednolity kod widziany przez redaktora. Przykładowy artykuł z jedną sekcją językową znajduje się na ilustracji 2.3.

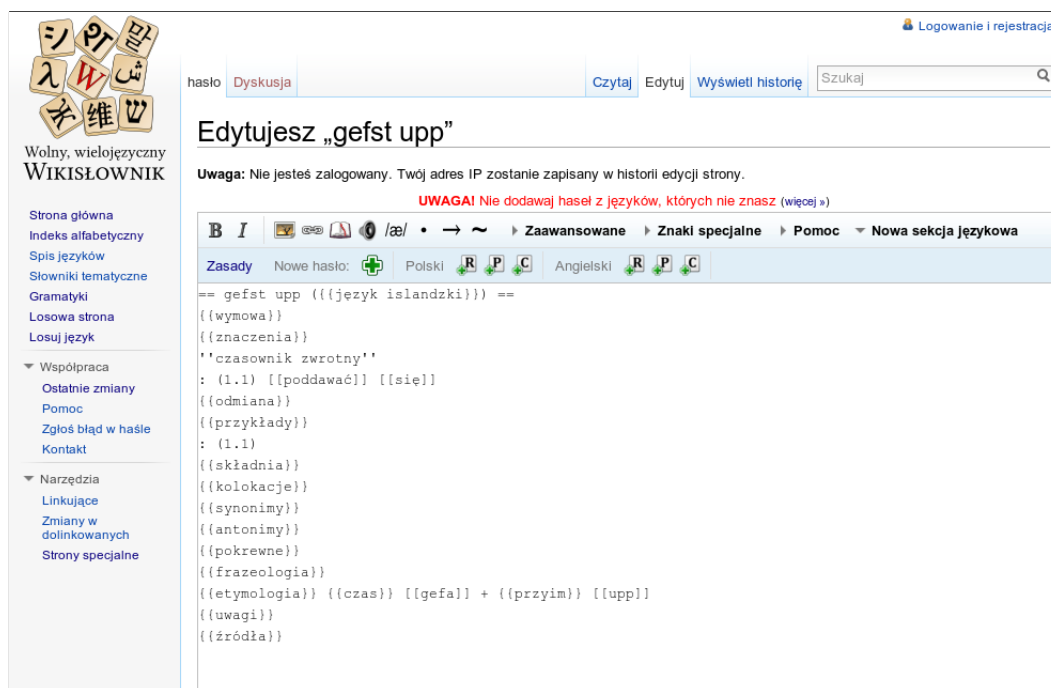


Ilustracja 2.3: Hasło *gefst upp* w polskim Wikisłowniku (http://pl.wiktionary.org/wiki/gefst_upp)

Na rzucie ekranu widoczny jest układ sekcji typowy dla polskiego projektu. Z małymi wyjątkami sekcje danego języka wyglądają tak samo we wszystkich hasłach. To oznacza, że także w innych opisanych wyrazach języka islandzkiego znaleźć można elementy: wymowa, znaczenia, odmiana, przykłady, składnia, kolokacje, synonimy, antonimy, wyrazy pokrewne, związki frazeologiczne, etymologia, uwagi, źródła. W rzeczywistości jest to układ typowy dla znakomitej większości występujących w Wikisłowniku języków [12]. Odstępstwa od tego szkieletu hasła są niewielkie i zostaną przedstawione poniżej.

Ilustracja 2.4 przedstawia wygląd przykładowej strony po przejściu do trybu edycji, dostępnego dla każdego użytkownika (bez konieczności logowania). Widok ten prezentuje

źródłowy wikikod hasła, dodatkowo edytor ma kilka ikon ułatwiających wstawianie często używanych elementów. W kodzie tym widoczne są szablony definiujące tytuły poszczególnych elementów sekcji (nazywanych dalej *podsekcjami*). Podział na sekcje i podsekcje jest dość prosty do wykonania za pomocą programu komputerowego i już sama ta procedura pozwala na znaczne zwiększenie przejrzystości w nowym edytorze, opisanym w rozdziale 4.



Ilustracja 2.4: Edycja hasła *gefst upp* w polskim Wikisłowniku (http://pl.wiktionary.org/w/index.php?title=gefst_upp&action=edit)

Każda z podsekcji w hasłach ma ściśle określoną funkcję, dzięki której polski Wikisłownik jest jednolity. Poniżej zostały omówione wszystkie podsekcje występujące w artykułach, w jedynej przyjętej jako poprawną kolejności. Duża ich część pojawia się jedynie w niektórych językach, dlatego choć lista podsekcji jest dość długa, to każde hasło będzie miało ich o wiele mniej. Szablony podsekcji zebrane są na stronie http://pl.wiktionary.org/wiki/Kategoria:Szablony_szablonów_haseł.

Podsekcje występujące w hasłach polskiego Wikisłownika

Szablon `{{zapis hieroglificzny}}`

Zawartość	Zapis hieroglificzny słowa w języku staroegipskim, pokazany za pomocą grafik PNG z repozytorium Wikimedia Commons. Oznaczenie (1.1) odnosi się do numeracji w sekcji <i>znaczenia</i> .
Języki	tylko staroegipski ² .
Przykład	<pre>{{zapis hieroglificzny}} : (1.1) [[Plik:Egyptian-Pr-cnḥ.PNG]]; [[Plik:Egyptian-Pr-cnḥ2.PNG]]; [[Plik:Egyptian-Pr-cnḥ3.PNG]] *</pre>
Szablon	{{ortografie}}
Zawartość	Inne sposoby zapisu tytułu hasła. Zazwyczaj chodzi o alternatywną pisownię w języku, do zapisu którego używane są dwa alfabety (np. serbski). Do prezentacji pisowni mogą być używane szablony wyświetlające dodatkowe informacje.
Języki	azerski, białoruski, dżuhuri, gagauski, krymskotatarski, ladino, serbski, slovio, tatarski, turkmeński, ujęurski
Przykład	<pre>{{ortografie}} Македониа *</pre>
Szablon	{{transliteracja}}
Zawartość	<p>Transliteracja słowa zapisanego w obcym alfabecie na alfabet łaciński.</p> <p>W przeciwieństwie do transkrypcji transliteracja może być wykonywana automatycznie – każda litera alfabetu obcego konwertowana jest na jeden lub więcej znaków w alfabecie łacińskim. Obecnie często używany jest szablon {{translit}}, który umożliwia automatyczną konwersję za pomocą JavaScriptu niektórych alfabetów podczas odczytywania strony.</p>
Języki	abazyński, abchaski, adygejski, akadyjski, amharski, arabski, aramejski, asamski, awarski, baszkirski, beludzi, bengali, birmański, bułgarski, chakański, czeczeński, czuwaski, dzongkha, erzja, gocki, gruziński, gudżarati, gyz, hebrajski, hindi, inguski, inuktitut, jidysz, kannada, kaszmirski, kazach-

²Rubryka *Języki* precyzuje, w których sekcjach językowych występuje dana podsekcja. Oprócz zwykłych sekcji, odpowiadających danemu językowi, istnieją także nietypowe: *użycie słowa obcego w języku polskim* i *znak chiński*

ski, khmerski, kirgiski, komi, komi-jażwiński, kri, kurdyjski, laotański, le-
zgiński, macedoński, malajalam, malediowski, marathi, maryjski, mongol-
ski, nepalski, newarski, nowogrecki, orija, ormiański, osetyjski, paszto, pen-
dżabski, perski, romski, rosyjski, sanskryt, sindhi, sorani, staro-cerkiew-
no-słowiański, starogrecki, staroormiański, sumeryjski, syngaleski, tabasa-
rański, tadżycki, tajski, tamazight, tamilski, telugu, tybetański, ukraiński,
urdu, zarfatit

Przykład `{{transliteracja}}` Moskva

*

Szablon `{{transkrypcja}}`

Zawartość Transkrypcja słowa zapisanego w obcym alfabecie na język polski, czyli przedstawienie go w formie dającej informacje o rzeczywistej wymowie

Języki standardowo tylko staroegipski, podsekcja ta może być jednak dodawana w wielu innych językach (przykład w jidysz)

Przykład `{{transkrypcja}}`
: (1.1-3) `{{YIV0|{{lp}} khaver {{lm}} khaveyrim}}`; polska: `{{lp}}`
chawer `{{lm}}` chawejrim
: (1.4) `{{YIV0|{{lp}} khover {{lm}} khovers}}`; polska: `{{lp}}`
chower; `{{lm}}` chowers

*

Szablon `{{czytania}}`

Zawartość Wyjaśnienie możliwego wymawiania znaków kanji używanych w języku japońskim. Występują dwa sposoby czytania: on’yomi i kun’yomi. Do ich prezentacji używane są szablony `{{on}}` i `{{kun}}`.

Języki tylko japoński

Przykład `{{czytania}}` `{{on}}` ビ (bi); `{{kun}}` はな (hana)

*

Szablon `{{klucz}}`

Zawartość Elementy, według których układane są słowniki języka chińskiego.

Języki tylko znak chiński

Przykład `{{klucz}}` 157 足 + 6

*

Szablon	<code>{{kreski}}</code>
Zawartość	Liczba kresek użytych do napisania danego znaku. Informacja służy m.in. do ułatwienia odnajdywania znaków w papierowych słownikach.
Języki	znak chiński, koreański
Przykład	<code>{{kreski}}</code> 13

*

Szablon	<code>{{warianty}}</code>
Zawartość	Szablon powiększający znak chiński i ewentualnie jego warianty. Szablon ten jest używany inaczej niż większość: nie odpowiada jedynie za wyświetlenie nagłówka, a zawartość sekcji wstawiana jest jako parametr. Najczęściej w parametrze pojawia się szablon <code>{{zch-w}}</code> , odpowiadający za prezentację znaku.
Języki	tylko znak chiński
Przykład	<code>{{warianty {{zch-w}}}}</code>

*

Szablon	<code>{{kolejność}}</code>
Zawartość	Kolejność stawiania kresek w znaku chińskim, ilustrowana za pomocą grafiki dostępnej w projekcie Wikimedia Commons. W tej podsekcji używane są szablony <code>{{zch-komiks}}</code> , <code>{{zch-cienie}}</code> i <code>{{zch-animacja}}</code> , które ładują automatycznie grafiki o nazwie odpowiadającej hasłu, jeśli te istnieją.
Języki	tylko znak chiński
Przykład	<code>{{kolejność}}</code> <code>{{zch-komiks}}</code>

*

Szablon	<code>{{wymowa}}</code>
Zawartość	Jedna z kluczowych podsekcji we wszystkich językach. Podawane są w niej informacje na temat wymowy danego hasła, zarówno za pomocą alfabetów fonetycznych (jak np. IPA oraz alfabet słowiański), jak i nagrań dźwiękowych umieszczonych w Wikimedia Commons. Do opisywania wymowy stosowane są szablony <code>{{IPA}}</code> , <code>{{IPA2}}</code> , <code>{{IPA3}}</code> , <code>{{IPA4}}</code> . Wymowa słów

polskich dodawana jest automatycznie przez bota uruchomionego przez jednego z administratorów Wikisłownika. Bot ten jest skomplikowanym programem napisanym w Javie [13], generującym wymowę na podstawie publikacji Danuty Ostaszewskiej i Jolanty Tambor [14]. Pliki dźwiękowe dodawane są szablonem `{{audio}}`.

Języki

wszystkie poza znakiem chińskim i użyciem międzynarodowym

Przykład

```
{{wymowa}} {{audio|Pl-samochód.ogg}}, {{IPA3|sã'mɔxut}},
{{AS3|sãm''o''χut}}, {{objaśnienie wymowy|WYG|NAZAL}}
```

*

Szablon

`{{znaczenia}}`

Zawartość

Jedyna podsekcja obowiązkowa, w której podawane jest znaczenie hasła. Zawartość podsekcji dzielona jest najpierw na części mowy, potem na poszczególne znaczenia, które zostają ponumerowane zgodnie z obowiązującym schematem. W hasłach polskich podawane jest dłuższe znaczenie, w innych językach tłumaczenie na polski. Znaczenia te są linkami do objaśnień form podstawowych poszczególnych słów – linkowanie to stanowi dość duże utrudnienie przy tworzeniu hasła.

Języki

wszystkie

Przykłady

Hasło angielskie:

```
{{znaczenia}}
''rzeczownik''
: (1.1) [[zamówienie]]
: (1.2) [[rozkaz]]
: (1.3) [[porządek]]
: (1.4) {{syst}} [[rząd]]
: (1.5) {{mat}} [[rząd]]
''czasownik''
: (2.1) [[zamawiać]]
: (2.2) [[rozkazywać]]
```

Hasło polskie:

```
{{znaczenia}}
```

```

''rzeczownik, rodzaj męski''
: (1.1) [[budynek]] [[warowny]]; {{wikipedia|zamek (architektura)}}
: (1.2) [[mechanizm]] [[zamykać|zamykający]] [[drzwi]],
[[szuflada|szuflady]]; {{wikipedia|zamek (urządzenie)}}
: (1.3) [[zapięcie]] [[garderoba|garderoby]],
{{zob|[[zamek błyskawiczny]]}}.
: (1.4) [[element]] [[składowy]] [[broń|broni]] [[palny|palnej]];
{{wikipedia|zamek (broń)}}
: (1.5) {{sport}} [[w]] [[hokej]]u: [[zamykać|zamknięcie]]
[[przeciwnik]]a [[w]] [[tercja|tercji]], [[gdy]] [[drużyna]]
[[atakować|atakująca]] [[rozgrywać|rozgrywa]] [[krążek]] [[w]]
[[tercja|tercji]] [[przeciwnik]]a [[nie]] [[pozwalać|pozwalając]]
[[on|mu]] [[wyjść]] [[poza]] [[niebieski|niebieską]] [[linia|linię]]

```

*

Szablon	{{determinatywy}}
Zawartość	Znak określający, o jaką klasę znaczeniową wyrazów chodzi w danym haśle. Wyświetlana jest grafika z Wikimedia Commons.
Języki	tylko staroegipski
Przykład	{{determinatywy}} <p>: (1.1) [[Plik:Egyptian-nb 'nh-determinative.PNG]]</p>

*

Szablon	{{odmiana}}
Zawartość	Odmiana wyrazu, prezentowana na różne sposoby. Występują np. szablony, które generują odmianę na podstawie formy podstawowej hasła. Niekiedy podawana jest jedynie deklinacja lub koniugacja (z odnośnikiem do tabel ją przedstawiających), gdzie indziej cała odmiana.
Języki	wszystkie poza znakiem chińskim
Przykład	{{odmiana}} <p>: (1.1-2) читать {{ter}} {{lp}} читаю, читаешь, читает; {{lm}} читаем, читаете, читают; {{przesz}} {{lp}} читал / читала / читало; {{lm}} читали; {{rozk}} {{lp}} читай; {{ims}} читающий; читаемый; читая</p>

*

Szablon	{{przykłady}}
Zawartość	Przykłady użycia danego słowa w zdaniu. W przypadku języków innych niż polski tłumaczenie przykładu podane jest po znaku →. Podobnie jak znaczenia, przykłady są linkowane.
Języki	wszystkie poza znakiem chińskim
Przykład	{{przykłady}} : (2.1) ''[[Michael]] ''locks'' [[his]] [[house]] [[every]] [[day]]. '' → [[Michał]] [[codziennie]] ''[[zamykać zamyka]]'' [[swój]] [[dom]] (na klucz).

*

Szablon	{{składnia}}
Zawartość	Podsekcja zawiera informacje o używaniu słowa w połączeniu z przyimkami czy przypadkami.
Języki	wszystkie poza znakiem chińskim
Przykład	{{składnia}} : (1.2) jechać +{{N}}; jechać [[do]] +{{D}}, jechać [[na]] +{{B}}

*

Szablon	{{kolokacje}}
Zawartość	Kolokacje to często używane zestawienia słów, w których (w przeciwieństwie do związków frazeologicznych) znaczenie całości wynika ze znaczenia poszczególnych wyrazów.
Języki	wszystkie poza znakiem chińskim
Przykład	{{kolokacje}} [[mieć]] / [[budzić]] / [[odbierać]] nadzieję • [[promyk]] nadziei • [[ziścić]] nadzieje • [[karmić]] [[się]] nadzieją

*

Szablon	{{synonimy}}
Zawartość	Wyrazy bliskoznaczne, synonimy.
Języki	wszystkie poza znakiem chińskim
Przykład	{{synonimy}} : (1.1) [[ufność]], [[wiara]], [[zawierzenie]], [[pociecha]]

*

Szablon	{{antonimy}}
Zawartość	Wyrazy przeciwstawne, antonimy
Języki	wszystkie poza znakiem chińskim
Przykład	{{antonimy}} : (1.1) [[rezygnacja]], [[zwątpienie]], [[beznadzieja]]

*

Szablon	{{złożenia}}
Zawartość	W językach japońskim i koreańskim podsekcja ta podaje słowa, które powstają jako złożenie danego hasła z innym. Użyty w przykładzie szablon {{furi}} pomaga dobrze wyświetlić furiganę – japońskie pismo.
Języki	koreański i japoński
Przykład	{{złożenia}} {{furi 五日 いつか}}, {{furi 五月 ごがつ}}, {{furi 五輪 ごりん}}, {{furi 五輪大会 ごりんたいかい}}

*

Szablon	{{pokrewne}}
Zawartość	Wyrazy pokrewne do danego wyrazu podstawowego. W przypadku większej grupy wyrazów wspólnej dla wielu haseł może wystąpić odsyłacz do wyrazu podstawowego, np. w hasle <i>kocur</i> : {{zob [[kot]]}}.
Języki	wszystkie poza znakiem chińskim
Przykład	{{pokrewne}} : (1.1) {{rzecz}} [[picklock]], [[locksmith]], [[locknut]] : (1.2) {{rzecz}} [[dreadlock]] : (1.3) {{rzecz}} [[airlock]], [[lockage]] : (2.1) {{przym}} [[lockable]]; {{rzecz}} [[locker]]

*

Szablon	{{pochodne}}
Zawartość	Odpowiednik podsekcji <i>pokrewne</i> dla morfemów w esperanto, stanowiących osobne hasła.
Języki	esperanto
Przykład	{{pochodne}} {{rzecz}} [[zebro]], [[zebrino]], [[zebrido]]

*

Szablon	{{frazologia}}
Zawartość	Podsekcja zawiera związki frazeologiczne, które prezentowane są podobnie jak kolokacje. Różnica między kolokacjami a związkami frazeologicznymi polega na tym, że w przypadku tych drugich znaczenie związku nie wynika bezpośrednio ze znaczeń poszczególnych wyrazów.
Języki	wszystkie poza znakiem chińskim
Przykład	<div>{{frazologia}}</div> <div>: [[psi urok]] • [[tu leży pies pogrzebany]] • {{wulg}} [[pies kogoś jebał]] • [[pies ogrodnika]] • {{pot}} [[pies na baby]] • [[pogoda pod psem]] • [[psu na budę]] • [[pieskie życie]] • [[psia wachta]] • [[psia koja]] • [[pies Pawłowa]] • [[nie dla psa kiełbasa]] • [[pies łańcuchowy Darwina]] • [[na psa urok]] • [[ni pies, ni wydra]] • [[schodzić na psy]] • [[łgać jak pies]] • [[delikatny jak francuski piesek]] • [[francuski piesek]] • [[pies z nim tańczył]] • [[psi żywot]] • [[psie figle]] • [[psi obowiązek]] • [[całować psa w nos]] • [[wyć jak pies do księżycy]] • [[żyć jak pies z kotem]] • [[wieszać na kimś psy]] • [[lubić kogoś jak psy dziada w ciasnej ulicy]] • [[robić coś psim śwędem]] • [[wyglądać jak zbity pies]] • [[być wyszczekany jak pies]] • [[psia kość]] • [[kupować za psie pieniądze]] • [[wyszczeakać coś jak pies]] • [[odszczekać coś jak pies]] • [[wierny jak pies]] • [[pies ci mordę lizał]]</div> <div>: zobacz też: [[Aneks:Przysłowia polskie - zwierzęta#pies przysłowia o psie]]</div>

*

Szablon	{{etymologia}}
Zawartość	Pochodzenie wyrazu, zapisywane za pomocą szablonów {{etym}} i {{etymn}}.
Języki	wszystkie
Przykład	<div>{{etymologia}}</div> <div>: {{etym prasłowiański *ne}} < {{etym praindoeuropejski *ne}} 'nie'</div> <div>: {{por}} {{etymn czeski ne}}, {{etymn rosyjski ne}},</div>

{{etymn|litewski|ne}}, {{etymn|łaciński|ne}}

*

Szablon {{kody}}

Zawartość Informacje na temat wprowadzania znaków chińskich za pomocą klawiatury w różnych metodach oraz kodowania Unicode. Podobnie jak w przypadku podsekcji *warianty*, i tutaj szablon przyjmuje parametry pozwalające na zestandaryzowane wyświetlanie.

Języki tylko znak chiński

Przykład {{kody |cjz= 田金 |cjl=WC |cr=6021₀ |u=56db}}

*

Szablon {{hanja}}

Zawartość W tej podsekcji podawana jest pisownia danego słowa koreańskiego w piśmie hanja (hancha), czyli pisownia zapożyczona z języka chińskiego. Częściej w koreańskim używany jest alfabet hangul.

Języki tylko koreański

Przykład {{hanja}} [[憲法]]

*

Szablon {{słowniki}}

Zawartość Informacja na temat występowania danego znaku chińskiego w słownikach KangXi, Dai Kanwa Jiten, Dae Jaweon i Hanyu Da Zidian.

Języki tylko znak chiński

Przykład {{słowniki|kx=1163.080|dkj=35533|dj=1628.020|hdz=63974.090}}

*

Szablon {{uwagi}}

Zawartość Dodatkowe informacje, np. częste błędy, odpowiedzi na typowe wątpliwości.

Języki wszystkie

Przykład {{uwagi}}

: (1.1) forma ''tylni'' dla przymiotnika rodzaju męskiego w liczbie pojedynczej jest błędna, może odnosić się ona jedynie do liczby mnogiej
<ref>{{PoradniaPWN|id=9687|hasło=tylny czy tylni?}}</ref>

*

Szablon	{{tłumaczenia}}
Zawartość	Podsekcja ta pełni funkcję słownika z języka polskiego na inne. Podawane są odnośniki do wyrazów będących tłumaczeniami danego słowa polskiego.
Języki	tylko polski
Przykład	<div><div>{{tłumaczenia}}</div><div><div><div>* angielski: (1.1) [[date]], [[appointment]]</div><div>* arabski: (1.1) [[تَعْيِين]]</div><div>* francuski: (1.1) [[rendez-vous]]</div><div>* rosyjski: (1.1) [[свидание]], {{pot}} [[свиданка]]</div><div>* szwedzki: (1.1) [[träff]] {{w}}</div></div></div></div>

*

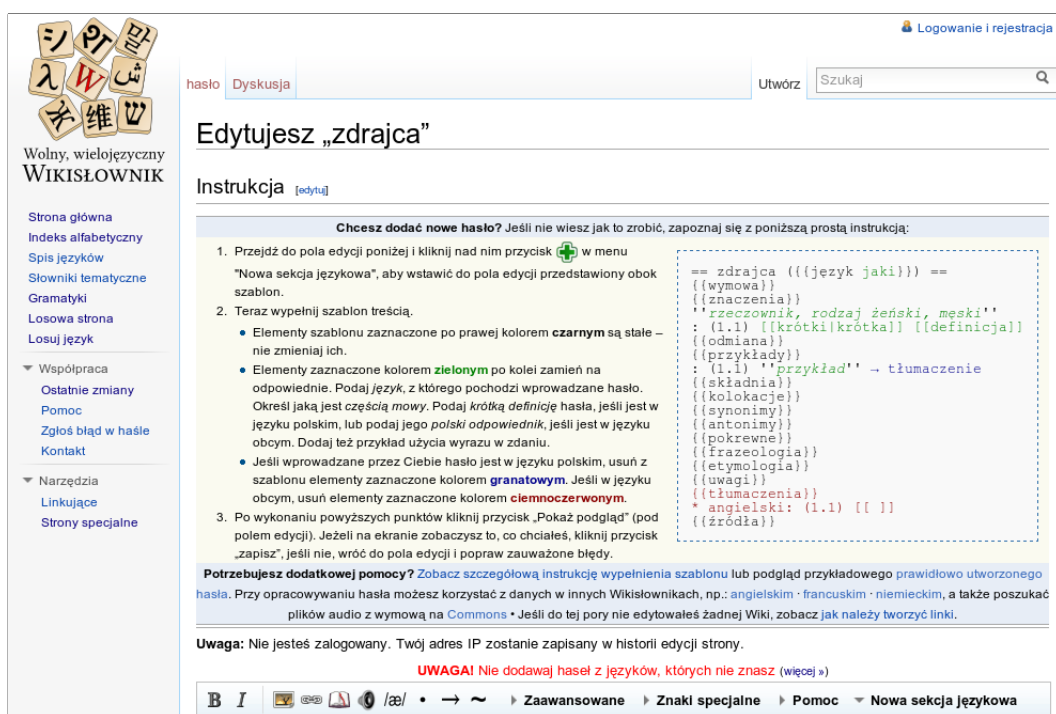
Szablon	{{źródła}}
Zawartość	Źródła dla informacji podanych w haśle. Zazwyczaj sekcja ta składa się ze znacznika <references/>, który powoduje wyświetlenie w tym miejscu przypisów wstawionych w poprzedzającej zawartości strony znaczników <ref>...</ref>.
Języki	wszystkie
Przykład	<div><div>{{źródła}}</div><div><references/></div></div>

2.4.2. Wady dotychczasowych rozwiązań

W dotychczasowym systemie edycji w Wikisłowniku przedstawiona wyżej regularna struktura nie jest wykorzystywana w wystarczającym stopniu. W dużej mierze wiąże się to z pierwotnym założeniem, jakim jest oparcie projektu na silniku MediaWiki. Każde hasło przechowywane jest w bazie danych jako jeden ciąg znaków, a ich struktura jest wymuszona dopiero przez społeczność – brakuje jakichkolwiek mechanizmów, które pomagałyby w utrzymywaniu haseł zgodnie ze standardem.

Baza Wikisłownika nie zachowuje nawet pierwszej postaci normalnej [15] – hasła nie są przecież danymi atomowymi. To duża wada projektu takiego jak słownik internetowy,

w którym dane powinny być prezentowane możliwie jednolicie. Postaci normalne pozwalają też na łatwiejszą modyfikację i obróbkę danych. Niestety niemożliwa jest zmiana silnika MediaWiki, dlatego należy dbać o standaryzację artykułów za pomocą innych środków. Aby pomieścić i logicznie powiązać informacje w każdym haśle, konieczne jest użycie omówionej struktury – sekcji i podsekcji, z zastosowaniem licznych szablonów. Edycja takiego hasła (p. ilustracja 2.4 na str. 16) będzie bardzo trudna dla początkującego redaktora Wikisłownika. Jeszcze większe problemy występują przy tworzeniu nowych haseł. Ilustracja 2.5 pokazuje ekran służący do tego celu (wywoływany np. po kliknięciu w czerwony link w dowolnym haśle, oznaczający nieistniejący do tej pory artykuł).



Ilustracja 2.5: Próba stworzenia nowego hasła w polskim Wikisłowniku

Jak widać na rzucie ekranu, sama instrukcja tworzenia nowego hasła jest na tyle skomplikowana, że może zniechęcić do edycji nowego redaktora. Wymagane jest skopiowanie podanego kodu i pieczołowite podmienianie poszczególnych jego elementów. Niestety tylko w bardzo ograniczonym stopniu uwzględniane jest, że niektóre języki mają zdecydowanie inną strukturę niż przykładowa. Osoba chcąca dodać do Wikisłownika nowy znak chiński będzie musiała nie tylko zapoznać się z regułami dla tej specyficznej odmiany haseł, ale też prawdopodobnie skopiować kod istniejącego hasła z tej kategorii. O wiele lepsza byłaby sytuacja, w której po wyborze języka edytujący otrzymuje odpowiedni szkielet hasła, który

będzie mógł uzupełnić standardowym formularzem, bez potrzeby wybierania fragmentów, które należy zmienić lub usunąć.

Oparcie projektów Wikimedia na systemie MediaWiki i wikikodzie powoduje tego typu problemy oczywiście nie tylko w Wikisłowniku. Także w Wikipedii komplikacje techniczne są dużą barierą dla nowych użytkowników, najczęściej przyzwyczajonych do narzędzi typu WYSIWYG (*What You See Is What You Get*) takich jak Microsoft Word. W badaniach przeprowadzonych w 2009 roku czytelnikom wszystkich wersji językowych Wikipedii zadano pytanie o powody nieuczestniczenia w pracach edycyjnych. 25,18% spośród 21 492 ankietowanych odpowiedziało *nie wiem, jak*, 12,49% stwierdziło, że zbyt słabo zna używane technologie, natomiast 24,43% wskazało, że obawiałoby się popełnienia błędu i „wpadnięcia w kłopoty” z tego powodu (najczęściej wymienianym powodem było *myślę, że nie mam wystarczającej wiedzy, którą mogę się podzielić* – 51,98%) [16]. Wynika z tego, że aż jedna czwarta czytelników, którzy wypełnili ankietę, to potencjalni redaktorzy, którzy nie angażują się w tworzenie encyklopedii ze względu na uwarunkowania techniczne. Brak podobnych badań przeprowadzanych na użytkownikach Wikisłownika, można jednak przypuszczać, że kształtują się one podobnie. Dodatkowym czynnikiem jest specjalizacja projektu – w pracach przy słowniku w dużej części uczestniczą lingwiści, którzy nie muszą biegle posługiwać się platformą MediaWiki.

W przypadku Wikipedii od dawna podejmowane są wysiłki w kierunku poprawienia obsługi technicznej serwisu. Istnieje np. alternatywny edytor haseł *wikEd* dostępny dla każdego użytkownika po wybraniu w preferencjach, w ograniczonym zakresie wspomagający redagowanie haseł [17]. Prawdopodobnie jednak edytor WYSIWYG wspólny dla przedsięwzięć Fundacji Wikimedia nie powstanie w przewidywalnej przyszłości. Z tego powodu korzystne jest usprawnianie poszczególnych projektów z uwzględnieniem ich cech charakterystycznych. Jednym z celów opisanej aplikacji jest zwiększenie komfortu korzystania z formularza wprowadzania haseł w Wikisłowniku. Implementacja tego projektu została omówiona w sekcji 4.2.

Drugą znaczącą wadą Wikisłownika jest wynikająca z wielojęzyczności ogromna redundancja, powodująca spore problemy dla czytelnika. Każda wersja językowa stanowi odrębną całość. Choć słowników internetowych nierzadko używają poliglotti, którym może być obojętne, w jakim języku zostaną zaprezentowane informacje, struktura Wikisłownika nie pozwala im wykorzystać swoich umiejętności. Specjalistyczne hasła mogą być dostępne

np. wyłącznie w jednej wersji językowej – przeszukiwanie poszczególnych edycji to marnotrawienie czasu, dlatego często najlepszym rozwiązaniem bywa po prostu użycie wyszukiwarki, np. Google. Informacje w różnych językach pojawiają się niezależnie od siebie, bowiem znakomita większość redaktorów jest aktywna wyłącznie w jednym projekcie. Przy tym zdarza się, że wprowadzający informacje wykonuje zupełnie niepotrzebną pracę, którą ktoś wykonał wcześniej w innej wersji językowej, a nawet w tej samej wersji, lecz podczas tworzenia innego hasła, w pewien sposób powiązanego z obecnie edytowanym.

Ważną funkcję we wszystkich projektach Wikimedia pełnią boty, odpowiadające w największych projektach za od 10 do 30% wszystkich edycji haseł [18]. W polskim Wikisłowniku programy te przede wszystkim uzupełniają linki interwiki, ale także wykonują bardziej skomplikowane akcje: standaryzują układ podsekcji i format linków, aktualizują indeksy i inne listy słów oraz dodają polską wymowę [19] [13]. Część z tych edycji jest wtórna i spowodowana błędami popełnianymi przez początkujących redaktorów – można by ich zatem uniknąć, usprawniając formularz edycji.

Automatyzacja edytora pozwala na sprawniejszą standaryzację i użycie informacji wprowadzonych już do innych wersji językowych Wikisłownika. Bardzo pomocne okazuje się tu API MediaWiki [20]. Ten interfejs programistyczny pozwala na sprawne pobieranie treści wybranych haseł z dowolnego projektu Wikimedia w celu dalszego przetwarzania. Zostało to wykorzystane podczas implementacji drugiej istotnej części aplikacji, a następnie opisane w sekcji 4.3.

2.4.3. Różnice w stosunku do Wikipedii

Po analizie uwidacznia się odmienny charakter Wikisłownika w stosunku do Wikipedii, która była pierwszym projektem Fundacji i dziś jest jej sztandarowym przedsięwzięciem. W tabeli 2.2 zostały ujęte najważniejsze różnice między witrynami w aspekcie technicznym.

To wszystko powoduje, że usprawnienie edycji w Wikisłowniku jest o wiele łatwiejsze niż w Wikipedii, której bardzo złożona struktura w praktyce uniemożliwia dziś kompleksową przebudowę procesu redakcji. Ściśle określona struktura to ogromna zaleta projektu. Co ciekawe – nie jest ona charakterystyczna dla wszystkich jego wersji językowych. Po części

Wikipedia	Wikisłownik
Struktura hasła różnorodna	Bardzo ściśle określona struktura hasła
Przed wszystkim ciągły tekst	Listy, wypunktowania
Umiarkowana liczba szablonów używanych w hasłach w stosunku do pozostałej treści	Bardzo duży udział szablonów w treści hasła
Duże różnice pomiędzy wersjami językowymi	Małe różnice pomiędzy wersjami językowymi
Linki interwiki łączą informacje o tym samym obiekcie w różnych językach	Linki interwiki łączą informacje o tym samym słowie występującym w wielu językach, objaśnione w różnych językach
Nagłówki dzielą artykuł na sekcje o dowolnej semantyce	Nagłówki dzielą artykuł na sekcje odpowiadające poszczególnym językom

Tabela 2.2: Porównanie Wikipedii i Wikisłownika – aspekty techniczne. Zob. też tabelę 3.1

prac włożonych w implementację okazało się, że polski Wikisłownik jest pod tym względem wyjątkowo dobrze zorganizowany. Wprawdzie wszystkie większe edycje zachowują standardowy podział na sekcje językowe, jednak owe sekcje niekiedy przybierają dość oryginalne postaci. Po dokładniejszych badaniach można stwierdzić, że polski Wikisłownik jest wyjątkowo jednolity, co ma niebagatelne znaczenie zarówno dla czytelnika, jak i dla edytora. W kolejnym rozdziale zanalizowano m.in. specyfikę Wikisłownika, jeśli skupić się na aspektach społecznościowych. Jak się okaże, także te elementy powinny mieć pozytywny wpływ na możliwość wprowadzenia usprawnień.

Rozdział 3

Aspekty społecznościowe

Tematem tego rozdziału są społecznościowe aspekty zarówno samego Wikisłownika, jak i niniejszej pracy magisterskiej. Na początku omówiono podstawowe założenia koncepcji *wiki*, na której opiera się Wikisłownik. Kolejna część rozdziału to krótka analiza społeczności rozwijającej projekt. Na dalszych stronach znalazły się treści ściśle związane z tworzoną aplikacją: opis rozwoju oprogramowania dla specyficznego klienta, jakim jest szersza społeczność, oraz szczegółowa analiza jego wymagań.

3.1. Koncepcja *wiki*

Pojęcie *wiki* określa witrynę internetową umożliwiającą tworzenie i edycję dowolnej liczby połączonych ze sobą hiperłączami stron przez przeglądarkę internetową, za pomocą uproszczonego języka znaczników lub edytora WYSIWYG [21]. Opisane w rozdziale 2 projekty Fundacji Wikimedia, szczególnie Wikipedia, są z pewnością najbardziej znanymi i największymi przykładami *wiki*.

Wiki to nie tylko technologia, ale także, a może przede wszystkim, społeczność. Rozmiary *wiki* rozciągają się od małych, zamkniętych witryn firmowych do największej Wikipedii. To, co łączy wszystkie tego typu projekty, to otwarty dostęp do edycji i tworzenia stron w obrębie danej grupy. W przypadku projektów Fundacji grupę tę stanowi niemal cała ludność świata dysponująca dostępem do internetu – poza nielicznymi krajami, w których sieć jest cenzurowana. Wśród redaktorów Wikipedii czy Wikisłownika nie istnieje żadna skodyfikowana hierarchia, co oznacza, że niezalogowany nowicjusz może zmienić stronę

dokładnie tak samo jak administrator projektu. Założenie to od lat stanowi główny zarzut pod adresem Wikipedii – negatywne opinie o tej encyklopedii są szeroko rozpowszechnione [22]. Mimo tego projekty stale rozwijają się i podnoszą swoją jakość – już w 2005 roku w czasopiśmie „Nature” opublikowano kontrowersyjne porównanie Wikipedii z *Encyclopædia Britannica*, z którego wynikało, że obie encyklopedie stoją na podobnym poziomie [23]. Jasno widać, że znaczenie projektów Fundacji Wikimedia jest dziś bardzo duże.

W tworzeniu Wikipedii i Wikisłownika biorą udział przedstawiciele całego społeczeństwa – w pracach uczestniczą zarówno profesorowie, jak i studenci. Jak wspomniano w sekcji 2.4.2, dla wielu z nich barierą jest poziom skomplikowania oprogramowania MediaWiki. Podczas gdy wielu uczestników, najczęściej tych o zainteresowaniach ścisłych, nie ma problemów z opanowaniem techniki edycji haseł, istnieje duża grupa osób, które deklarują, że mogłyby zaangażować się bardziej, gdyby nie konieczność opanowania złożonych sposobów wprowadzania informacji. Dzięki regularnej strukturze akurat polski Wikisłownik jest tym projektem, w którym uproszczenie tworzenia i edycji haseł jest najbardziej możliwe do przeprowadzenia. Ponieważ proces taki wymaga wspólnej decyzji ogółu społeczności, należało zanalizować, jak kształtują się interakcje między użytkownikami w projekcie.

3.2. Społeczność polskiej edycji Wikisłownika

W podsekcji 2.4.3 przedstawione zostały różnice między technicznymi elementami polskich wersji Wikipedii i Wikisłownika. Uporządkowanie haseł w Wikisłowniku daje nadzieję na wprowadzenie ulepszeń – główną barierą staje się więc społeczność, która stara się podejmować decyzje konsensualnie.

Najlepszym sposobem przedstawienia społeczności Wikisłownika jest przeciwstawienie jej grupy osób edytujących w Wikipedii¹. W tabeli 3.1 przedstawiono najbardziej widoczne różnice między nimi, które można zauważyć już po kilkugodzinnej lekturze archiwalnych dyskusji i historii edycji.

Można uznać, że w przypadku polskich wersji językowych Wikisłownik jest projektem znacznie spokojniejszym niż Wikipedia, której powszechność powoduje nieuniknione problemy, takie jak zaangażowanie osób chcących wykorzystać projekt do celów marketingowych lub ideologicznych czy liczne wandalizmy w wykonaniu znudzonych nastolatków.

¹Od tej pory pod pojęciami *Wikipedia* i *Wikisłownik* domyślnie będą kryć się wersje polskojęzyczne.

Wikipedia	Wikisłownik
Dużo wandalizmów (edycji wykonywanych w złej wierze)	Bardzo mało wandalizmów
Kilkuset stale aktywnych redaktorów	Kilkunastu stale aktywnych redaktorów
Wielu użytkowników mających kłopoty z dostosowaniem się do zasad	Niewielu użytkowników mających kłopoty z dostosowaniem się do zasad
Liczne wojny edycyjne (wzajemne cofanie swoich edycji przez co najmniej dwóch redaktorów)	Praktycznie brak wojen edycyjnych
Niezliczone strony archiwalnych dyskusji	Łatwy dostęp do archiwalnych dyskusji
Dobra dokumentacja i pomoc	Dobra pomoc, ale szczegółowa dokumentacja bardzo uboga
Bardzo zróżnicowane obszary zainteresowania	Społeczność o wspólnych zainteresowaniach lingwistycznych

Tabela 3.1: Porównanie Wikipedii i Wikisłownika – aspekty społecznościowe. Zob. też tabelę 2.2

Społeczność redaktorów słownika zdecydowanie skupiona jest na stałym udoskonalaniu projektu. Z racji jego rozmiaru łatwiej jest o konsensus w dyskusjach, a dzięki mniejszej popularności innowacyjne rozwiązania techniczne mają większą szansę na realizację.

3.3. Analiza wymagań

3.3.1. Wymagania funkcjonalne

3.3.2. Wymagania nefunkcjonalne

3.3.3. Spodziewane korzyści

3.4. Specyfika tworzenia aplikacji dla wikispołeczności

Rozdział 4

Opis implementacji

W ostatnim rozdziale szczegółowo przedstawiony została implementacja aplikacji. Opisano użyte technologie i metodologię, a szczególną uwagę poświęcono problemom, które powstały podczas procesu tworzenia programu. Pierwsza sekcja zawiera opis środowiska, w którym osadzona jest aplikacja, natomiast kolejne dwie skupiają się na dwóch logicznie wyodrębnionych częściach projektu: uproszczonym formularzu edycyjnym oraz automatyzacji tworzenia haseł. W ostatniej części rozdziału przedstawiono przebieg wdrożenia aplikacji w polskim Wikisłowniku i możliwości dalszego rozwoju.

4.1. Wprowadzenie

Oprogramowanie MediaWiki zostało napisane w języku PHP (*PHP Hypertext Preprocessor*) i jest wysoce konfigurowalne za pomocą tzw. rozszerzeń, dodających kolejne funkcje – przykładowo: obsługę przypisów, dodatkowe strony specjalne czy nowe funkcje parsera [24]. W projektach Fundacji obsługa rozszerzeń jest kontrolowana przez Fundację. Poszczególne projekty mogą składać prośby o włączenie określonego rozszerzenia, decyzję podejmują zaś główni programiści Fundacji. Teoretycznie można rozważać utworzenie aplikacji dla Wikisłownika jako rozszerzenia w PHP. Opcja ta została jednak odrzucona we wstępnej fazie projektu – stworzenie rozszerzenia wymagałoby nieporównanie więcej formalności niż aplikacja kliencka, przede wszystkim jednak jego funkcjonalność byłaby znacznie ograniczona: poza polskim Wikisłownikiem żaden inny projekt nie mógłby z niego skorzystać, na pewno nie zostałby więc włączony do głównej wersji MediaWiki.

Podobnie jak w większości przypadków w projektach Wikimedia jako metodę dostosowania silnika MediaWiki do szczególnych potrzeb wybrano zatem aplikację napisaną w JavaScript. Konieczne jest zatem przedstawienie sposobu obsługi skryptów na stronach tych witryn. Pliki JavaScript ładowane są z wielu różnych źródeł – hierarchia w nieco uproszczonej postaci jest następująca [25]:

1. skrypty systemowe oprogramowania MediaWiki, wspólne dla wszystkich projektów oraz charakterystyczne dla danego projektu ze względu na ładowane rozszerzenia,
2. skrypt zapisany na poziomie danego projektu jako strona `MediaWiki:Common.js` – dostęp do niego mają lokalni administratorzy,
3. skrypt zapisany na poziomie danego projektu dla wybranej przez użytkownika skórki (domyślną jest *Vector*, poprzednią był *Monobook*) jako strona, np. `MediaWiki:Vector.js` – dostęp do niego mają lokalni administratorzy,
4. tzw. gadżety, czyli dodatkowe skrypty rozszerzające funkcjonalność, możliwe do włączenia w preferencjach użytkownika – także one edytowane są przez administratorów,
5. skrypt zapisany na stronie danego użytkownika, np. `User:Sokrates/common.js` – na tym poziomie możliwe jest dostosowywanie skryptów do swoich potrzeb przez każdego użytkownika,
6. skrypt zapisany na stronie danego użytkownika dla wybranej skórki, np. `User:Sokrates/vector.js`.

Jak widać, możliwości dołączania dodatkowych skryptów są elastyczne i dostępne na różnych poziomach dla różnych użytkowników. Decyzja o włączeniu skryptów dla wszystkich użytkowników danego projektu należy do jego administratorów (nie licząc wspólnych skryptów narzuconych przez Fundację), natomiast każdy użytkownik poprzez edycję specjalnej strony może uruchamiać różne skrypty na swoje potrzeby, nie uciekając się do konieczności korzystania z dodatków do przeglądarek tj. Greasemonkey.

4.2. Formularz edycyjny

4.3. Automatyzacja edycji hasła

4.4. Wdrożenie i dalszy rozwój

Rozdział 5

Podsumowanie

Bibliografia

- [1] Stowarzyszenie Wikimedia Polska. Fundacja Wikimedia. http://pl.wikimedia.org/w/index.php?title=Fundacja_Wikimedia&oldid=10461. [Dostęp 6 sierpnia 2011].
- [2] List of Wikipedias – Meta, discussion about Wikimedia projects. http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=2715916. [Dostęp 6 sierpnia 2011].
- [3] Wikimedia Foundation. <http://www.wikimedia.org/>. [Dostęp 5 sierpnia 2011].
- [4] Licensing update/Result – Meta, discussion about Wikimedia projects. http://meta.wikimedia.org/w/index.php?title=Licensing_update/Result&oldid=1776040, 2009. [Dostęp 5 sierpnia 2011].
- [5] Help:Contents – MediaWiki, The Free Wiki Engine. <http://www.mediawiki.org/w/index.php?title=Help:Contents&oldid=421744>, 2011. [Dostęp 5 sierpnia 2011].
- [6] Dlaczego Wikisłownik? – Wikisłownik, wolny wielojęzyczny słownik. http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Dla_czego_Wikis%C5%82ownik&oldid=2292379, 2007. [Dostęp 5 sierpnia 2011].
- [7] Administrators – Wiktionary. <http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Administratorzy&oldid=1316722>, 2003. [Dostęp 5 sierpnia 2011].
- [8] Administratorzy – Wikisłownik, wolny wielojęzyczny słownik. <http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Administratorzy&oldid=1316722>, 2004. [Dostęp 5 sierpnia 2011].
- [9] Administrators of Wikimedia projects/Wikipedias – Meta, discussion about Wikimedia projects. http://meta.wikimedia.org/w/index.php?title=Administrators_of_Wikimedia_projects/Wikipedias&oldid=2748948, 2011. Dostęp 5 sierpnia 2011.

- [10] Strona główna – Wikisłownik, wolny wielojęzyczny słownik. http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Strona_g%C5%82%C3%B3wna&oldid=2298705, 2004. [Dostęp 5 sierpnia 2011].
- [11] Wiktionary: List of Wiktionaries – Meta, discussion about Wikimedia projects. http://meta.wikimedia.org/w/index.php?title=Wiktionary&oldid=2424480#List_of_Wiktionaries. [Dostęp 5 sierpnia 2011].
- [12] Zasady tworzenia haseł – Wikisłownik, wolny wielojęzyczny słownik. http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Zasady_tworzenia_hase%C5%82&oldid=2294171, 2004. [Dostęp 5 sierpnia 2011].
- [13] O. Matyja. Wikipedysta:Olafbot – Wikisłownik, wolny wielojęzyczny słownik. <http://pl.wiktionary.org/w/index.php?title=Wikipedysta:Olafbot&oldid=2123349>. [Dostęp 6 sierpnia 2011].
- [14] D. Ostaszewska, J. Tambor. *Fonetyka i fonologia współczesnego języka polskiego*. Wydawnictwo Naukowe PWN, Warszawa, wydanie 2, dodruk 4, sierpień 2009.
- [15] C.J. Date. *An Introduction to Database Systems*. Addison-Wesley, Boston, wydanie 8, 2004.
- [16] R. Glott, Ph. Schmidt, R. Ghosh. Wikipedia Survey – Overview of Results. Raport instytutowy, UNU-MERIT, Maastricht, 15 marca 2010. Dostępny w sieci: http://www.wikipediastudy.org/docs/Wikipedia_Overview_15March2010-FINAL.pdf [Dostęp 5 sierpnia 2011].
- [17] User:Cacycle/wikEd – Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org/w/index.php?title=User:Cacycle/wikEd&oldid=442519985>, 2011. [Dostęp 5 sierpnia 2011].
- [18] R. S. Geiger. The lives of bots. *Critical Point of View. A Wikipedia Reader*, strony 78–93. Institute of Network Cultures, Amsterdam, 2011. Dostępny w sieci: http://www.networkcultures.org/_uploads/%237reader_Wikipedia.pdf [Dostęp 6 sierpnia 2011].
- [19] Wikisłownik:Boty – Wikisłownik, wolny wielojęzyczny słownik. <http://pl.wiktionary.org/w/index.php?title=Wikis%C5%82ownik:Boty&oldid=2020877>. [Dostęp 6 sierpnia 2011].
- [20] API:Main Page – MediaWiki, The Free Wiki Engine. http://www.mediawiki.org/w/index.php?title=API:Main_page&oldid=407852. [Dostęp 6 sierpnia 2011].

- [21] M. A. Dennis. *wiki*. Encyclopædia Britannica. Encyclopædia Britannica Online. <http://www.britannica.com/EBchecked/topic/1192819/wiki>, 2011. [Dostęp 27 sierpnia 2011].
- [22] S. Vaknin. Wikipedia, Criticism of. <http://knol.google.com/k/wikipedia-criticism-of>, 2008. [Dostęp 27 sierpnia 2011].
- [23] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438:900–901, 2005. doi: 10.1038/438900a.
- [24] Manual:Extensions – MediaWiki, The Free Wiki Engine. <http://www.mediawiki.org/w/index.php?title=Manual:Extensions&oldid=429069>. [Dostęp 28 sierpnia 2011].
- [25] Wikipedia:Technik/Skin/JS – Wikipedia. <http://de.wikipedia.org/w/index.php?title=Wikipedia:Technik/Skin/JS&oldid=92936675>. [Dostęp 28 sierpnia 2011].

Spis ilustracji

2.1	Polska edycja Wikipedii	8
2.2	Polska edycja Wikisłownika	12
2.3	Hasło w polskim Wikisłowniku	15
2.4	Edycja hasła w polskim Wikisłowniku	16
2.5	Próba stworzenia nowego hasła w polskim Wikisłowniku	27

Spis tabel

2.1	Porównanie HTML i wikitekstu	11
2.2	Porównanie Wikipedii i Wikisłownika – aspekty techniczne	30
3.1	Porównanie Wikipedii i Wikisłownika – aspekty społecznościowe	33