

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

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 łącz 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

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.3 Informatyka

Klasyfikacja tematyczna

D. Software

D.2. Software Engineering

D.2.10. Design

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	9
2.3. Wiktory – Wikisłownik	11
2.4. Polska edycja Wikisłownika	13
2.4.1. Struktura hasła	13
3. Aspekty społecznościowe	25
3.1. Koncepcja <i>wiki</i>	25
3.2. Społeczność polskiej edycji Wikisłownika	25
3.3. Specyfika tworzenia aplikacji dla wikispołeczności	25
4. Opis implementacji	27
4.1. Wprowadzenie	27
4.2. Formularz edycyjny	27
4.3. Automatyzacja edycji hasła	27
4.4. Wdrożenie	27
5. Podsumowanie	29
Bibliografia	29

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 korzystaliby, 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 *Encyclopaedia 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

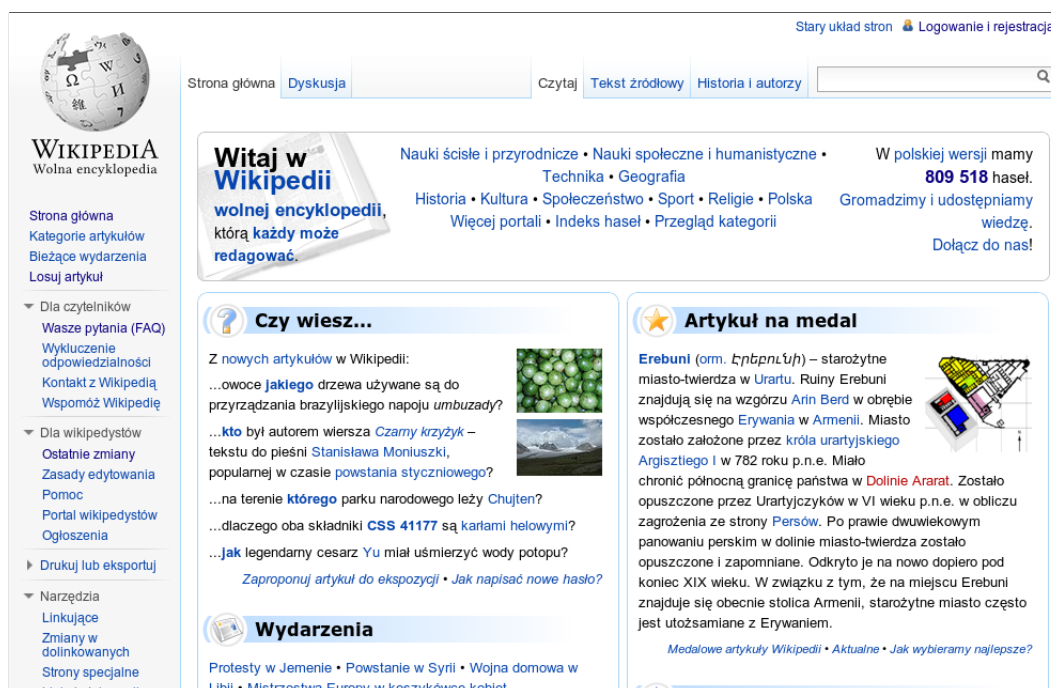
Rozdział ten stanowi charakterystykę Wikisłownika – sieciowego słownika opartego na oprogramowaniu MediaWiki. Wikisłownik jest jednym z największych i najpopularniejszych słowników dostępnych w polskim internecie. W kolejnych sekcjach projekt ten został opisany na różnych poziomach szczegółowości. Omówiono 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*. Doskonale znaną, sztandarową inicjatywą Fundacji jest Wikipedia (<http://www.wikipedia.org>) – największa obecnie encyklopedia internetowa, dostępna w 281 językach (stan z maja 2011 roku) i zawierająca ponad 18 milionów haseł, w tym ponad 3,6 miliona w największej, angielskojęzycznej¹ edycji. Mimo częstej krytyki tego 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.

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

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



Ilustracja 2.1: Polska edycja Wikipedii

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

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). 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 wolontarystyczna 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 two-

rzyć za ich pomocą stron bez uprzedniego dłuższego przygotowania. Aby umożliwić bezproblemową edycję stron internetowych osobom bez wykształcenia informatycznego, programiści MediaWiki zaprojektowali tzw. wikitekst – uproszczony język opisu stron, pozwalający na realizację wymienionych elementów. Porównanie niektórych z nich znajduje się w tabeli 2.1.

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

Tabela 2.1: Porównanie HTML i wikitekstu

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

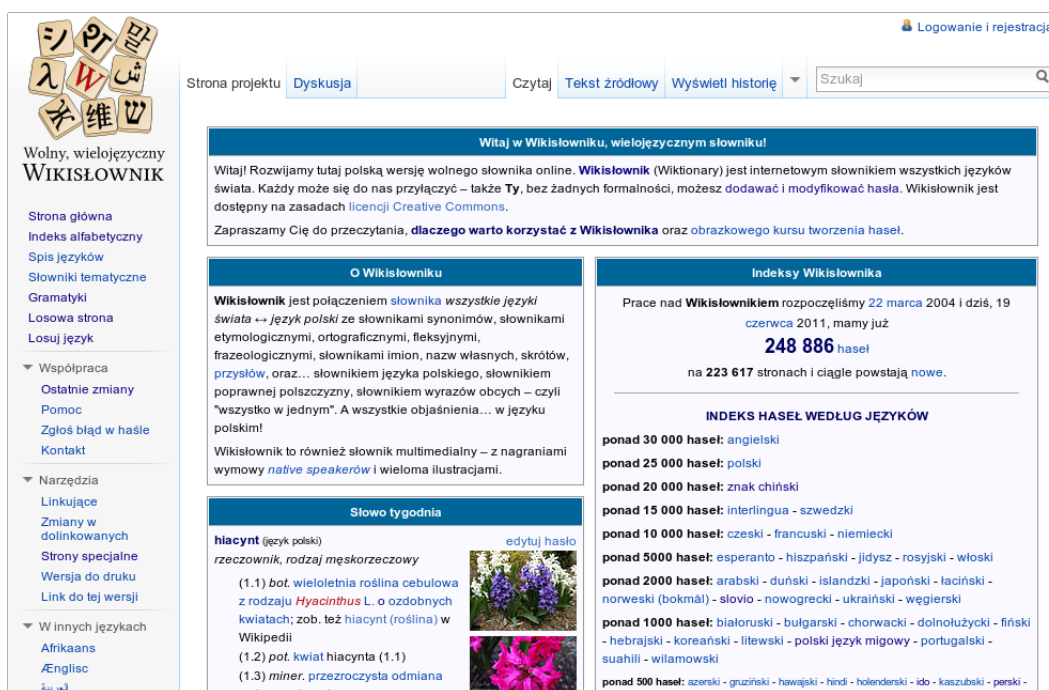
Bardzo istotnym elementem wikitekstu są szablony – predefiniowane fragmenty kodu z opcjonalnymi parametrami. Szablony można uznać za odpowiednik procedur/funkcji w językach programowania. W projektach opartych na MediaWiki szablony pełnią przede wszystkim dwie główne funkcje:

- 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=wartość}}. 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 Wiktionary homepage. At the top left is the Wiktionary logo with the text 'Wolny, wielojęzyczny WIKISŁOWNIK'. Below it are navigation links: 'Strona główna', 'Indeks alfabetyczny', 'Spis języków', 'Słowniki tematyczne', 'Gramatyki', 'Losowa strona', 'Losuj język'. A 'Współpraca' section includes 'Ostatnie zmiany', 'Pomoc', 'Zgłoś błąd w haśle', and 'Kontakt'. A 'Narzędzia' section includes 'Linkujące', 'Zmiany w dolinkowanych', 'Strony specjalne', 'Wersja do druku', and 'Link do tej wersji'. A 'W innych językach' section lists 'Afrikaans', 'English', and 'عربية'. The main content area has a search bar and a 'Logowanie i rejestracja' link. Below the search bar is a welcome message: 'Witaj w Wikisłowniku, wielojęzycznym słowniku!'. It explains that Wiktionary is an online Polish version of a free dictionary and that users can add or modify entries. It also mentions the Creative Commons license. Below this is a section 'O Wikisłowniku' which describes the project's goals and the types of entries it includes. To the right is a section 'Indeksy Wikisłownika' showing the number of entries and pages in various languages. At the bottom is a section 'Słowo tygodnia' featuring the word 'hiacynt' (hyacinth) with its definition and images of the flowers.

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.

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, zaś w polskim – 14. Dla porównania Wikipedia w języku angielskim ma 1541 administratorów (niekoniecznie aktywnych), wersja polska natomiast 163.

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 (patrz 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 (patrz sekcja 2.4). Ze względu na odmiennność poszczególnych wersji językowych prawdopodobnie w innych nie będzie można wykorzystać 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. 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.

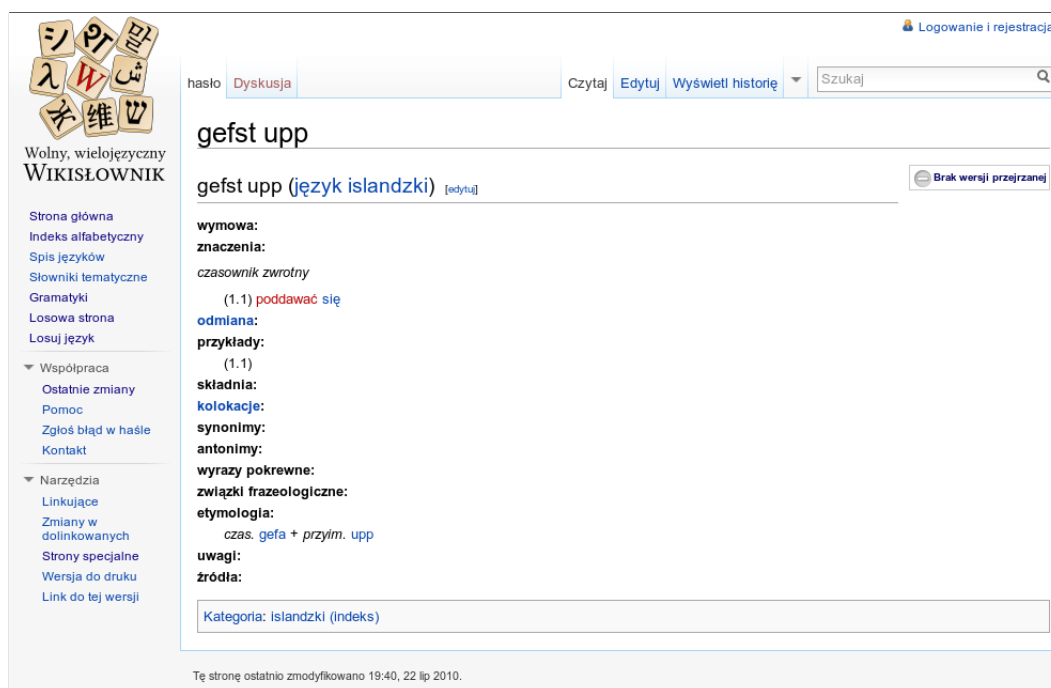
W kolejnej podsekcji opisana została dokładnie struktura haseł w polskim Wikisłowniku. Społeczność tego projektu scharakteryzowano natomiast w sekcji 3.2.

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.

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. 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



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

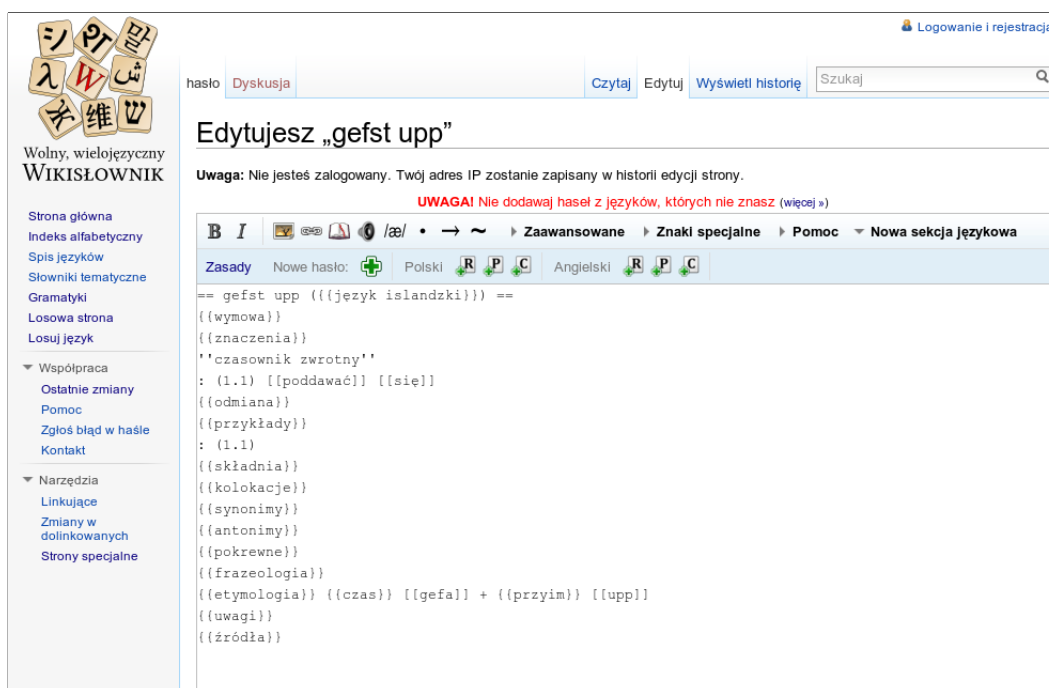
widoczne są szablony definiujące tytuły poszczególnych elementów sekcji (nazywanych dalej *podsekcjami*). Podział na sekcje i podsekcje jest dość prosty do wykonania za pomocą programu komputerowego i już sama ta procedura pozwala na znaczne zwiększenie przejrzystości w nowym edytorze, opisanym w rozdziale 4.

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	<code>{{zapis hieroglificzny}}</code>
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	<code>{{zapis hieroglificzny}}</code> : (1.1) [[Plik:Egyptian-Pr-cnḥ.PNG]];

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*



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

[[Plik:Egyptian-Pr-cnḥ2.PNG]];

[[Plik:Egyptian-Pr-cnḥ3.PNG]]

*

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 serbski, gagauski, dżuhuri, ladino, tatarski, krymskotatarski, turkmeński, ujugurski, slovio, azerski, białoruski

Przykład {{ortografie}} Македониа

*

Szablon {{transliteracja}}

Zawartość Transliteracja słowa zapisanego w obcym alfabecie na alfabet łaciński. 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.

Języki rosyjski, ukraiński, kazachski, ormiański, gruziński, hebrajski, jidysz, arabski, perski, tajski, urdu, hindi, bułgarski, macedoński, starogrecki, nowogrecki, tybetański, abazyński,

abchaski, adygejski, akadyjski, amharski, aramejski, assamski, bengali, awarski, baszkirski, beludzi, birmański, gyyz, gudżarati, gocki, erzja, kirgiski, dzongkha, czuwaski, czeceński, chakaski, bośniacki, tamilski, inguski, inuktitut, kannada, kaszmirski, marathi, khmerski, komi, komi-jażwiński, kri, kurdyjski, laotański, lezgiński, malajalam, malediowski, mongolski, nepalski, newarski, osetyjski, paszto, pendżabski, romski, sanskryt, sindhi, sorani, staro-cerkiewno-słowiański, sumeryjski, syngaleski, tabasarański, tadżycki, tamazight, telugu, 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 {{kreski}}

Zawartość Liczba kresek użytych do napisania danego znaku. Informacja służy m.in. do ułatwienia odnajdywania znaków w papierowych słownikach.

Języki znak chiński, koreański

Przykład `{{kreski}}` 13

*

Szablon `{{warianty}}`

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 `{{zch-w}}`, odpowiadający za prezentację znaku.

Języki tylko znak chiński

Przykład `{{warianty|{{zch-w}}}}`

*

Szablon `{{kolejność}}`

Zawartość Kolejność stawiania kresek w znaku chińskim, ilustrowana za pomocą grafiki dostępnej w projekcie Wikimedia Commons. W tej podsekcji używane są szablony `{{zch-komiks}}`, `{{zch-cienie}}` i `{{zch-animacja}}`, które ładują automatycznie grafiki o nazwie odpowiadającej hasłu, jeśli te istnieją.

Języki tylko znak chiński

Przykład `{{kolejność}}`
`{{zch-komiks}}`

*

Szablon `{{wymowa}}`

Zawartość Jedna z kluczowych podsekcji we wszystkich językach. Podawane są w niej informacje na temat wymowy danego hasła, zarówno za pomocą alfabetów fonetycznych (IPA oraz alfabet słowiański), jak i nagrań dźwiękowych umieszczonych w Wikimedia Commons. Do opisywania wymowy stosowane są szablony `{{IPA}}`, `{{IPA2}}`, `{{IPA3}}`, `{{IPA4}}`. Wymowa słów polskich dodawana jest automatycznie przez bota uruchomionego przez jednego z administratorów Wikisłownika. Bot ten jest skomplikowanym programem napisanym w Javie, generującym wymowę na podstawie publikacji Ostaszewskiej i Tambora. Pliki dźwiękowe dodawane są szablonem `{{audio}}`.

Języki wszystkie poza znakiem chińskim i użyciem międzynarodowym

Przykład `{{wymowa}}` `{{audio|PL-samochód.ogg}}`, `{{IPA3|sǎ'mɔxuːt}}`, `{{AS3|sǎm''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 hasle. Wyświetlana jest grafika z Wikimedia Commons.

Języki

tylko staroegipski

Przykład

```
{{determinatywy}}
: (1.1) [[Plik:Egyptian-nb 'nh-determinative.PNG]]
```

*

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}} : (1.1-2) читать {{ter}} {{lp}} читаю, читаешь, читает; {{lm}} читаем, читаете, читают; {{przesz}} {{lp}} читал / читала / читало; {{lm}} читали; {{roz}} {{lp}} читай; {{ims}} читающий; читаемый; читая

*

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]]'' [[sвій]] [[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	{{frazologia}} : [[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 swędem]] • [[wyglądać jak zбитy 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]] : zobacz też: [[Aneks:Przysłowia polskie - zwierzęta#pies przysłowia o psie]]

*

Szablon	{{etymologia}}
Zawartość	Pochodzenie wyrazu, zapisywane za pomocą szablonów {{etym}} i {{etymn}}.
Języki	wszystkie
Przykład	{{etymologia}} : {{etym prasłowiański *ne}} < {{etym praındoeuropejski *ne}} 'nie' : {{por}} {{etymn czeski ne}}, {{etymn rosyjski ne}}, {{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 <i>warianty</i> , 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 (han-cha), 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}} <p>: (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></p>

*

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	{{tłumaczenia}} <p>* angielski: (1.1) [[date]], [[appointment]]</p> <p>* arabski: (1.1) [[تَعْيِين]]</p> <p>* francuski: (1.1) [[rendez-vous]]</p> <p>* rosyjski: (1.1) [[свидание]], {{pot}} [[свиданка]]</p> <p>* szwedzki: (1.1) [[träff]] {{w}}</p>

*

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	{{źródła}}
	<references/>

*

Rozdział 3

Aspekty społecznościowe

3.1. Koncepcja *wiki*

3.2. Społeczność polskiej edycji Wikisłownika

3.3. Specyfika tworzenia aplikacji dla wikispołeczności

Rozdział 4

Opis implementacji

4.1. Wprowadzenie

4.2. Formularz edycyjny

4.3. Automatyzacja edycji hasła

4.4. Wdrożenie

Rozdział 5

Podsumowanie