

PRZEDMIOT: Metody Inżynierii Wiedzy – Projekt  
TYTUŁ: Moduł obsługi baz danych BibTeX dla  
DokuWiki

Grzegorz Miklaszewski `grmiklas@student.agh.edu.pl`  
AiR, IV rok, lato 2007

13 czerwca 2007

---

## Spis treści

<b>1</b>	<b>Sformułowanie problematyki projektu</b>	<b>3</b>
<b>2</b>	<b>Cel projektu</b>	<b>4</b>
<b>3</b>	<b>Opis realizacji</b>	<b>4</b>
3.1	Koncepcja . . . . .	4
3.2	Wybrane narzędzia . . . . .	5
3.3	Realizacja . . . . .	6
<b>4</b>	<b>Podsumowanie</b>	<b>6</b>
4.1	Wnioski . . . . .	6
4.2	Instrukcja obsługi . . . . .	7
4.2.1	Instalacja . . . . .	7
4.2.2	Składnia . . . . .	7

## 1. Sformułowanie problematyki projektu

DokuWiki to oprogramowanie do szybkiego i prostego tworzenia dokumentacji w postaci strony www. Napisane w języku PHP charakteryzuje się brakiem wymagań odnośnie ingerencji w kod źródłowy ze strony użytkownika. Nie jest także potrzebna baza danych. Użytkownik posługuje się systemem za pomocą wbudowanego interfejsu graficznego, pozwalającego na zaawansowaną rozbudowę strony. Dodatkowo istnieje możliwość rozszerzenia funkcjonalności DokuWiki o tzw. wtyczki, czyli dodatkowe komponenty dołączane do systemu.

**[[pl:miw:miw2007\_tematy]]** AIWiki

Pokaż źródło strony | Poprzednie wersje | Ostatnie zmiany | Szukaj

Ślad: » miw2007\_tematy  
Jesteś tutaj: start » pl » miw » miw2007\_tematy

### Uwagi do Studentów

- Tematy oznaczone znacznikiem należy uznać za tematy o wysokim dla Autorów priorytecie, a co za tym idzie autorzy będą silniej je wspierać i
- Tematy oznaczone znacznikiem należy uznać za tematy o podwyższonym priorytecie
- Tematy pozostałe też są ważne i też można dostać ocenę pozytywną, w ty, 5.0 i 6.0, są one jednak potencjalnie mniejszym wyzwaniem, lub mniej pilnie potrzebne autorom

### Kontynuacja tematów

Osoby, które zrealizują w sposób wyróżniający się, a także będą zainteresowane tematyką poruszaną w projektach, będą mogły kontynuować pracę w ramach tematyki pochodnej do MIW 2007 w ramach:

- pracy magisterskiej u dr. inż. G. J. Nalepy lub u dr. inż. I. Wojnickiego, *uaktualniona* lista tematów będzie dostępna na wiosnę w systemie
- projektu

### Tematy MIW07

#### RBS

*RBS\_RuleBurst*  
Analiza narzędzia, pakietu, podejścia. <http://www.ruleburst.com>

*RBS\_EBRC*  
Przegląd tematyki konferencji, sponsorów, narzędzi, postaci. <http://www.eurobizrules.org>

*RBS\_JREAPI*  
standaryzacja mechanizmów regułowych w Javie, stan na dziś, na ile, co i jak implementują narzędzia, np. Jess <http://jcp.org/en/jsr/detail?id=94>

Rysunek 1: Przykładowy dokument DokuWiki

BibTeX jest narzędziem służącym do formatowania danych bibliograficznych. Posiada własną składnię opisu informacji w postaci tekstowej. Sama baza tylko przechowuje dane, nie umożliwia ich wyświetlenia ani przetworzenia na inną formę. Do tego służą zewnętrzne programy potrafiące odczytać pliki BibTeX. Przykładem jest system do składania tekstu  $\text{\LaTeX}$ . Informacje przechowywane w plikach z rozszerzeniem .bib mogą być dołączane do dowolnego dokumentu  $\text{\LaTeX}$  bez konieczności każdorazowego ręcznego uzupełniania w nim bibliografii. Dodatkowo istnieje wiele szablonów wyświetlania danych BibTeX, np. APA, MLA i Chicago.

W standardowej dystrybucji DokuWiki nie było możliwości korzystania z bazy BibTeX. Nie posiadała ona funkcji rozpoznawania składni BibTeX w edytorze ani wczytania bazy z pliku. Jedyny sposób na utworzenie bibliografii polegał na ręcznym przepisaniu pozycji z bazy do formatu zgodnego z DokuWiki. Istniała wtyczka umożliwiająca rozpoznawanie składni BibTeX w edytorze DokuWiki, jednak generowała kod niezgodny z językiem XHTML. Dodatkowo istniała jedynie możliwość ręcznego uzupełniania wpisów, co przy większych ilościach danych było bardzo niewygodne.

```

@ARTICLE
{
  klitzing:nobel,
  AUTHOR={Klaus von Klitzing},
  TITLE="The Quantized Hall Effect", JOURNAL=RMP,
  VOLUME=58,
  PAGES=519,
  YEAR=1986
}

```

Rysunek 2: Przykładowy zapis w formacie BibTeX

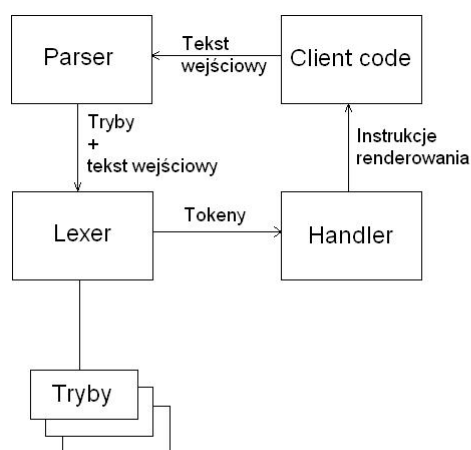
## 2. Cel projektu

Projekt miał na celu stworzenie rozszerzenia do systemu DokuWiki umożliwiającego korzystanie z formatu BibTeX. W ten sposób, po dodaniu wtyczki użytkownik miałby możliwość skopiować zawartość pliku BibTeX do edytora DokuWiki. System sam przekształciłby tekst na odpowiedni format wyjściowy. Efektem końcowym byłaby wygenerowana bibliografia, zgodna z odpowiednim formatem, zapisana w języku XHTML.

## 3. Opis realizacji

### 3.1. Koncepcja

Do budowy wtyczki został użyty standardowy mechanizm DokuWiki odpowiedzialny za przekształcanie tekstu, wpisanego do edytora, na format XHTML. Aby system DokuWiki mógł wykryć obecność składni BibTeX, należało umieścić ją wewnątrz znaczników `<BIBTEX></BIBTEX>`. Pomiedzy nimi nie była rozpoznawana żadna inna składnia niż BibTeX. Na rysunku przedstawiono schemat generowania kodu dokumentu DokuWiki:




Rysunek 3: Diagram relacji między komponentami

1. Lexer skanuje dokument DokuWiki w poszukiwaniu ciągów znaków odpowiadających znanym schematom i zwraca sekwencje podzielonych ciągów znaków na wyjściu
2. Handler otrzymuje tokeny (partie podzielonych ciągów znaków) od modułu Lexer i przekształca je w sekwencje instrukcji zrozumiałych dla części Renderer
3. Parser stanowi połączenie pomiędzy modułami Lexer oraz Handler dostarczając im reguły formatowania dla treści DokuWiki
4. Renderer przekształca instrukcje otrzymane od modułu Handler i przekształca je na kod XHTML

Ogólne etapy realizacji rozszerzenia do systemu DokuWiki wyglądały następująco:

- Do modułu Parser został dodany prosty szablon umożliwiający wykrycie znaczników BibTeX podczas skanowania tekstu przez moduł Lexer
- Moduł Handler otrzymał funkcję pozwalającą przetworzyć otrzymane ciągi znaków na sekwencje instrukcji.
- W module Renderer została umieszczona najważniejsza część implementacji. Została do niego dodana funkcja korzystająca z pakietu OSBib, odpowiedzialna za generowanie kodu XHTML.

Efektom końcowym była bibliografia jak na poniższym rysunku:

```
[1980, article]
von Klitzing, K., Dorda, G., & Pepper, M. (1980). New method for high accuracy determination of fine structure
constant based on quantised hall resistance. Phys. Rev. Lett., 45, 494.
[1986, article]
von Klitzing, K. (1986). The Quantised Hall Effect. Rev. Mod. Phys., 58, 519.
[2007, misc |  www]
Wikipedia. (2007). Mouse --- Wikipedia{,} The Free Encyclopedia.
```

Rysunek 4: Przykład bibliografii na stronie www

### 3.2. Wybrane narzędzia

Projekt został zrealizowany w oparciu o bibliotekę OSBib. Zapewniła ona wszelkich koniecznych funkcji potrzebnych do parsowania kodu Bibtex. Wszelkie informacje o bibliotece OSBib są dostępne na stronie <http://bibliophile.sourceforge.net/downloads.php>. Pakiet ten został stworzony przez Mark Grimshaw, Andrea Rossato, Guillaume Gardey, Christian Boulanger i jest rozpowszechniany na zasadzie licencji GPL. Celem użycia tej biblioteki w projekcie była zamiana długiego ciągu znaków na tablicę z uporządkowaną strukturą, posortowaną zgodnie z poszczególnymi atrybutami formatu BibTeX. Na podstawie tej tablicy, poprzez odczytanie jej poszczególnych elementów, został wygenerowany kod XHTML. Do utworzenia wtyczki skorzystano z języka PHP. Było to jedyne możliwe rozwiązanie, ponieważ zarówno powyższa biblioteka jak i samo DokuWiki są napisane w tym języku.

### 3.3. Realizacja

W poszczególnych plikach pakietu OSBib oraz wtyczki Bibtex zostały dokonane zmiany:

- style.css Kod napisany od nowa
- syntax.php
  - function handle(\$match, \$state, \$pos)  
Rozbudowanie funkcji o możliwość rozpoznawania pochodzenia wpisów Bibtex. Dodany kod sprawdzał jaki tryb wpisów został użyty ("url", "text", "file"). Na końcu funkcja dołącza odczytaną informację do zwracanej tablicy.
  - function render(\$mode, &\$amp;renderer, \$data)  
Do funkcji *createCitations* jest przekazywany dodatkowy parametr z informacją opisaną powyżej.
  - function createCitations(&\$data, &\$mode)  
Dodanie argumentu odbierającego wartość przekazaną od funkcji *render*. Przypisanie stałej, ścieżki do plików wysyłanych przez użytkowników do DokuWiki. W tej funkcji znajduje się istotny fragment kodu odpowiedzialnego za odczytywanie nazw plików ze ścieżki lub adresu url. Korzysta z mechanizmu wyrażeń regularnych. Dodatkowo przeprowadzono kilka drobnych zmian w sposobie generowania kodu HTML. W tym celu podmieniono część znaczników, aby kod był bardziej zgodny z XHTML.
- /OSBib/format/bibtexParse/PARSEENTRIES.php
  - function openBib(\$file)  
Przy przesyłaniu ścieżki do pliku, funkcja mająca na celu rozpoznać jej poprawność zwracała błąd jeśli jest to adres URL. Została ujęta w komentarz.

## 4. Podsumowanie

### 4.1. Wnioski

Nie wszystkie założenia projektu udało się spełnić. Jednym z nich była zgodność kodu ze specyfikacją XHTML. Okazało się to niemożliwe ze względu na sposób wstawiania nieformatowanych bloków tekstowych. DokuWiki standardowo otacza je znakiem paragrafu `<p>`. Zgodnie ze standardem XHTML wewnątrz paragrafu nie mogą się znajdować znaczniki definicji, a takie są konieczne.

Zauważono problemy z obsługą dużych i złożonych plików BIB. Przyczyna znajduje się przypuszczalnie w bibliotece OSBib. Błędu nie udało się naprawić.

Kolejną cechą, której nie udało się zaimplementować, była obsługa stylów nagłówków pozycji bibliograficznych. Problematiczne było np. kolejne numerowanie pozycji. Ponieważ funkcja tworząca kod wyjściowy jest uruchamiana dla każdego załączanego pliku oddzielnie, nie miała możliwości odczytania czy kolejny plik jest częścią poprzedniej bibliografii czy zupełnie nową, numerowaną od początku. PHP nie udostępnia prostej możliwości sprawdzenia tego faktu.

Ostatni błąd polegał na braku wcięć pozycji bibliograficznych. Nie udało się określić przyczyny, ale wszystko wskazywało na to, że tkwiła ona w stylach CSS DokuWiki. Po ich wyłączeniu błąd zniknął. Bardzo prawdopodobne jest, że po wgraniu wtyczki do innej wersji DokuWiki błąd nie wystąpi.

Poza powyższymi niedogodnościami, wtyczka działa poprawnie. Istnieje możliwość dołączania plików zarówno na serwerze lokalnym jak i z zewnątrz. Użytkownika ma także możliwość dodania wpisów ręcznie. Poprawiono i uproszczono generowany kod HTML, zmieniając także style CSS.

## 4.2. Instrukcja obsługi

### 4.2.1. Instalacja

Aby zainstalować wtyczkę, należy przenieść cały folder z zawartością do katalogu */lib/plugins/* wewnątrz DokuWiki. Przed użyciem należy upewnić się, że *allow\_url\_open* w PHP jest włączone. W przeciwnym wypadku dołączanie plików poprzez adresy url spowoduje błąd.

### 4.2.2. Składnia

Istnieją trzy możliwości dodawania wpisów Bibtex:

**Pliki z zewnątrz** Pliki spoza serwera lokalnego dodaje się podając ich adres www jako odnośnik zewnętrzny DokuWiki. Składnia jest następująca:

```
<bibtex mode="url">[[adres url]]</bibtex>
```

**Pliki lokalne** Pliki wysłane przy użyciu mechanizmu DokuWiki dodaje się jako odnośnik wewnętrzny. Składnia jest następująca:

```
<bibtex mode="file">{{nazwa pliku}}</bibtex>
```

**Bezpośrednie wpisy** Istnieje możliwość ręcznego wpisania pozycji Bibtex:

```
<bibtex mode="text">
```

```
wpisy Bibtex
```

```
</bibtex>
```