

Jak stworzyć filtry umożliwiające indeksowanie wybranych typów dokumentów w SAP NetWeaver Portal?

Może zaistnieć sytuacja, że chcemy zindeksować jakieś repozytorium danych, po to by mieć możliwość pełno-tekstowego wyszukiwania informacji. Jednak w tym repozytorium zgromadzone są dokumenty zapisane w różnych formatach, często w takich, z których bardzo rzadko, a nawet w ogóle nie korzystamy. Chcąc przyspieszyć proces indeksowania danych, oraz zawęzić późniejsze wyniki wyszukiwania tylko do dokumentów zapisanych w wybranych przez nas formatach danych należy stworzyć odpowiednie filtry.

Result Resource Filter jest filtrem, którego głównym zadaniem jest ograniczenie działania tzw. „crawler(ów)” wykorzystywanych w procesie indeksowania danych. Najczęściej wykorzystywanymi formatami danych w przedsiębiorstwach są dokumenty zapisane w popularnych formatach takich jak np.: *.doc, *.ppt, *.pdf, *.xls, *.html i praktycznie tylko te dokumenty są przez pracowników wykorzystywane. Jeżeli zatem nie istnieje potrzeba wyszukiwania informacji zapisanych w innych formatach danych, możemy skorzystać z wspomnianego powyżej filtra, by zawęzić wyniki wyszukiwania do tych formatów danych.

Procedura definiowania Result Resource Filter:

1. Idź do *System Administration* → *System Configuration* → *Knowledge Management* → *Content Management* → *Global Services* → *Resource Filters* → *Result Resource Filter*.

2. Wybierz *New*

Mamy tutaj 2 możliwości. Albo możemy podać formaty dokumentów, które chcemy aby zostały zindeksowane lub też możemy podać formaty dokumentów, których nie chcemy indeksować. Różnica jest taka, że w pierwszym przypadku zindeksowane zostaną dokumenty tylko określonego typu np.: *.doc, *.ppt, *.pdf, natomiast w drugim zostaną zindeksowane wszystkie dokumenty poza tymi, które wymieniliśmy jako te, które nie mają być brane pod uwagę np.: *.zip, *.xml itp.

3. Wybierając pierwszą opcję z rozwijanego menu *Item ID Mode* wybierz *include*.

4. W pole *Item ID Patterns* wpisz odpowiednie formaty danych (np. *.ppt, *.doc, *.pdf

Name *	filtr_1
Include Documents/Web-Pages	<input checked="" type="checkbox"/>
Include Folders	<input checked="" type="checkbox"/>
Include Links (Not Applicable For Web-Sites)	<input checked="" type="checkbox"/>
Case Sensitive (Folders And Documents/Web-Pages)	<input type="checkbox"/>
Item ID Mode (Documents/Web-Pages Only) *	include
Item ID Patterns (csv)	*.ppt, *.doc, *.pdf
Mime Type Mode (Documents/Web-Pages Only) *	exclude
Mime Type Patterns (csv)	
Minimum Content Size (Documents/Web-Pages Only)	0 Bytes
Maximum Content Size (Documents/Web-Pages Only)	0 Bytes
Maximum Age of Last Modification (Documents/Web-Pages Only)	0 Days

OK Apply Cancel

- Wybierając drugą opcję z rozwijanego menu *Item ID Mode* wybierz *exclude*.
- W pole *Item ID Patterns* wpisz odpowiednie formaty danych (np. *.zip, *.xml).

New Result Resource Filter

Name *	filtr_2
Include Documents/Web-Pages	<input checked="" type="checkbox"/>
Include Folders	<input checked="" type="checkbox"/>
Include Links (Not Applicable For Web-Sites)	<input checked="" type="checkbox"/>
Case Sensitive (Folders And Documents/Web-Pages)	<input type="checkbox"/>
Item ID Mode (Documents/Web-Pages Only) *	exclude
Item ID Patterns (csv)	*.zip, *.xml
Mime Type Mode (Documents/Web-Pages Only) *	exclude
Mime Type Patterns (csv)	
Minimum Content Size (Documents/Web-Pages Only)	Bytes
Maximum Content Size (Documents/Web-Pages Only)	Bytes
Maximum Age of Last Modification (Documents/Web-Pages Only)	Days

OK Apply Cancel

Jeżeli chcesz aby crawler rozpoznawał duże i małe litery należy zaznaczyć opcję *Case Sensitive*.

- Wybierz OK.
- Tworząc teraz nowy profil crawler(a) do wyboru będą stworzone powyżej filtry.



Autor:
Paweł Pawlas
Konsultant Zespołu e-business

Tekst przygotowany w marcu 2006.
Wszystkie nazwy i znaki użyte w tekście są znakami zastrzeżonymi.