

Parsing

Ten tryb służy do przetworzenia plików zebranych w trybie *Crawling* do jednolitej dla wszystkich źródeł struktury danych. Parsowane dane również są umieszczane w repozytorium w postaci pliku tekstowego w formacie JSON. Schemat struktury danych dostępny jest [tutaj](#).

Dodatkowe parametry:

Salon24

Dziel dane na bloki - [**Tak/Nie**] parametr określający czy parsowane dane mają być dzielone na bloki. To źródło danych jest bardzo duże i przy parsowaniu dużej jego części może dojść do sytuacji gdy zabraknie pamięci operacyjnej, aby zapisać dane do repozytorium.

Maksymalny rozmiar bloku posta - maksymalna ilość postów w jednym bloku. Parametr ma zastosowanie jeśli i tylko jeśli parametr *Dziel dane na bloki* ma wartość TAK.

Maksymalny rozmiar bloku komentarza - maksymalna ilość komentarzy w jednym bloku. Parametr ma zastosowanie jeśli i tylko jeśli parametr *Dziel dane na bloki* ma wartość TAK.

Zbieraj statusy - [**TAK/NIE**] parametr określa czy parsowane posty mają mieć zbierane również dane o statusie społecznym (statusy w mediach społecznościowych Facebook, Twitter, Google+). Wymaga to jednak dodatkowego łączenia się z tymi serwisami, co znacznie spowalnia pracę usługi.

Podgląd uruchomionych zadań

Po uruchomieniu zadania, możemy sprawdzić jego status przechodząc do listy zadań w zakładce menu górnego [Zadania] -> [Moje zadania]

Complex Networks | Zadania | GridFTP | Repozytorium | Anotator | Marcin Kulisiewicz

Lista zadań

Wyszukaj frazę

Moje zadania
Zleć zadanie QCG
Edytor QCG
Zleć syndykacje

< 1 > z 10

Opis	Wysłane	Start	Koniec	Status	Host	Typ
Zadanie syndykacji	11 paź 2015, 21:16			QUEUED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 12:01	8 paź 2015, 12:01	8 paź 2015, 12:01	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 12:00	8 paź 2015, 12:00	8 paź 2015, 12:07	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:54	8 paź 2015, 09:54	8 paź 2015, 09:54	FAILED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:51	8 paź 2015, 09:51	8 paź 2015, 09:51	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:46	8 paź 2015, 09:46	8 paź 2015, 09:46	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:44	8 paź 2015, 09:44	8 paź 2015, 09:44	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:42	8 paź 2015, 09:42	8 paź 2015, 09:42	FINISHED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:36	8 paź 2015, 09:36	8 paź 2015, 09:36	FAILED		syndication szczegóły
Zadanie syndykacji	8 paź 2015, 09:34	8 paź 2015, 09:34	8 paź 2015, 09:35	FAILED		syndication szczegóły

1 2 3 4 5 z 10 [następna](#)

Bezpośrednio po uruchomieniu zadanie będzie posiadało status **QUEUED**.

Statusy zadań:

- **QUEUED** - zadanie czeka w kolejce na uruchomienie,
- **RUNNING** - zadanie jest w trakcie wykonywania,
- **FAILED** - zadanie zakończyło się błędem w trakcie wykonywania,
- **FINISHED** - zadanie zakończyło się poprawnie.

Jeżeli zadanie ma status **FINISHED**, możliwe jest pobranie wyników zadania. Aby pobrać wyniki należy wybrać w górnym menu pozycję [Repozytorium](#) i przejść do Community *Complex Networks* i kolekcji *Syndykacja*. Więcej informacji w [rozdziale podręcznika opisującym Repozytorium](#).

Zaawansowane użycie

Aby skorzystać z adnotowania danych patrz [podręcznik użytkownika AnnotationHelper](#).