

SPECIFICA TECNICA

Il presente documento si riferisce alla fornitura di dati bibliometrici organizzati in un db mysql, costruito secondo quanto segue.

Fonte di riferimento: Web of Science – *Core Collection*, ossia:

- Science Citation Index Expanded
- Social Sciences Citation Index
- Arts & Humanities Citation Index
- Conference Proceedings Citation Index- Science + Social Science & Humanities
- Book Citation Index– Science + Social Sciences & Humanities
- Emerging Sources Citation Index (ESCI)

Periodo di riferimento: 2001-2017

Modalità di estrazione:

- parola “Italy” nell’affiliation di almeno un autore delle pubblicazioni indicizzate nella fonte di riferimento
- document type ammissibili: Art Exhibit Review; Article; Book; Book Chapter; Book Review; Chronology; Dance Performance Review; Database Review; Discussion; Excerpt; Fiction, Creative Prose; Film Review; Hardware Review; Letter; Music Performance Review; Music Score; Music Score Review; Poetry; Proceedings Paper; Record Review; Review; Script; Software Review; Theater Review; TV Review, Radio Review; Data Paper

Indicizzazione dei dati finali: per autore accademico “disambiguato”, ossia ricercatori (RTD e RTI) e professori (di prima e seconda fascia), in ruolo per almeno un anno nel periodo in osservazione, presso una università italiana.

A quest’ultimo proposito, la “disambiguazione” delle authorship prevede che ciascuna pubblicazione venga attribuita ai ricercatori/docenti accademici che l’hanno realizzata, tramite un complesso algoritmo a sua volta basato:

- sull’accoppiamento dei dati estratti da WoS con i record del database CINECA¹ sul personale di ruolo degli Atenei italiani al 31/12 di ciascun anno;
- su un successivo intervento di filtraggio necessario a individuare e rimuovere possibili false attribuzioni generate da problemi di diversa natura, in particolare di omonimia².

Modalità di consegna: il db dovrà essere costituito al minimo da tre tabelle:

- *Authorship*: tracciato a due campi, ID docente e ID pubblicazione
- *Docenti*: per ogni docente accademico italiano, oltre ad un ID univoco, occorre indicare tutte le anagrafiche disponibili su sito CINECA¹
- *Pubblicazioni*: per ogni record pertinente della WoS core collection di Clarivate, oltre ad un ID univoco, occorre indicare tutti i metadati bibliografici estraibili da tale repertorio.

¹ <http://cercauniversita.cineca.it/php5/docenti/cerca.php>

² La media armonica di *precision* e *recall* (F-measure) delle authorships disambiguate dall’algoritmo è pari a circa il 97% (valore calcolato su base campionaria, con errore di campionamento del 2% e un livello di confidenza del 98%).