

Metodo di strutturazione e compressione di alberi etichettati aventi grado e forma arbitrari



INVENTORI: Ferragina Paolo
Manzini Giovanni
Muthukrishnan S.
Luccio Fabrizio

STATUS PATENT: concesso

N° PRIORITÀ: US11783288

DATA DI CONCESSIONE: 10/04/2012

L'invenzione



L'invenzione brevettata riguarda un **metodo per comprimere alberi etichettati e supportare ricerche efficienti su di essi.**

L'etichettatura degli alberi è sui nodi, e la tecnica brevettata consente di ottenere prestazioni in compressione prossime all'entropia di ordine-k calcolata sul contenuto e sulla struttura dell'albero.

Le operazioni che possono essere eseguite sull'albero compresso la navigazione della sua struttura, oppure ricerche di sequenze di etichette che occorrono come cammini nell'albero. Ciascuna di queste operazioni può essere eseguita efficientemente senza decomprimere l'albero interamente, ma agendo solo su una piccola porzione di esso.

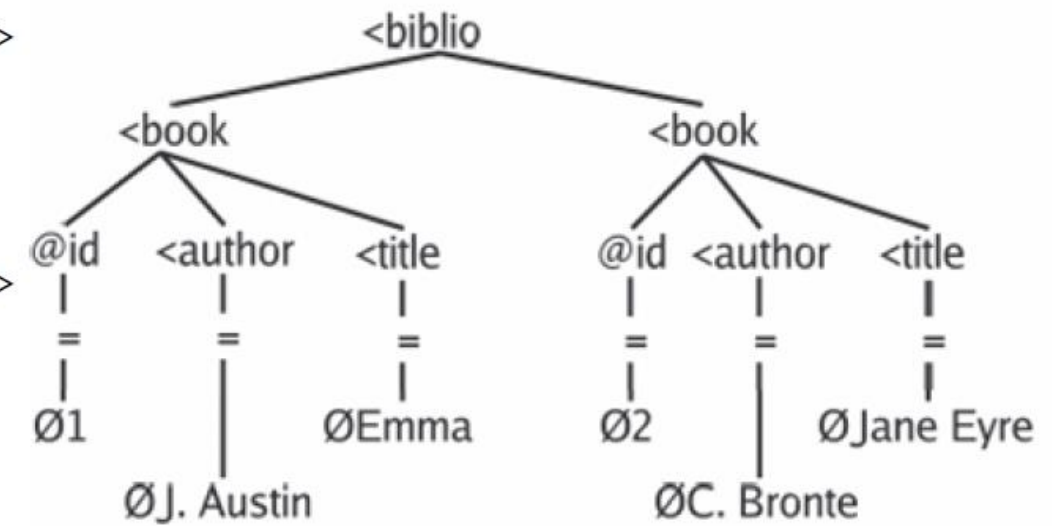
L'applicazione più naturale è la compressione e ricerca in **documenti in formato XML o JSON.**

Il risultato è stato oggetto di alcune pubblicazioni scientifiche apparse su conferenze (ACM-SIAM Symp. on Discrete Algorithms e WWW Conference) e riviste (Journal of the ACM) prestigiose che ne hanno dimostrato anche la sua efficienza pratica.

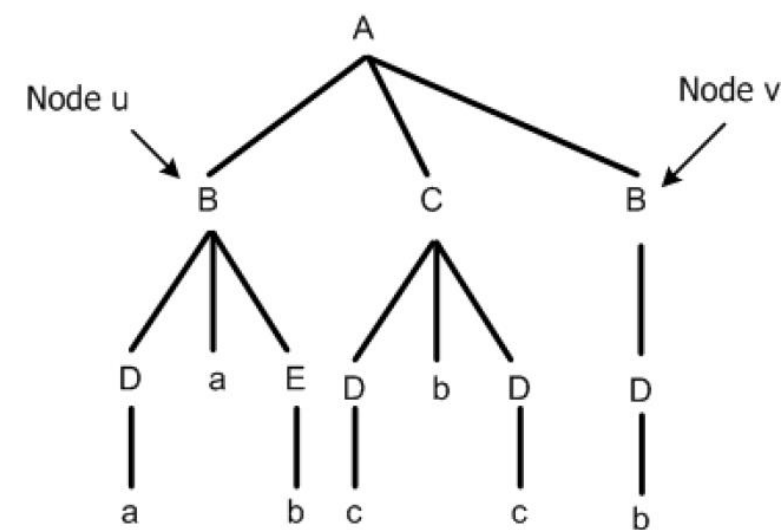


Un esempio di albero etichettato

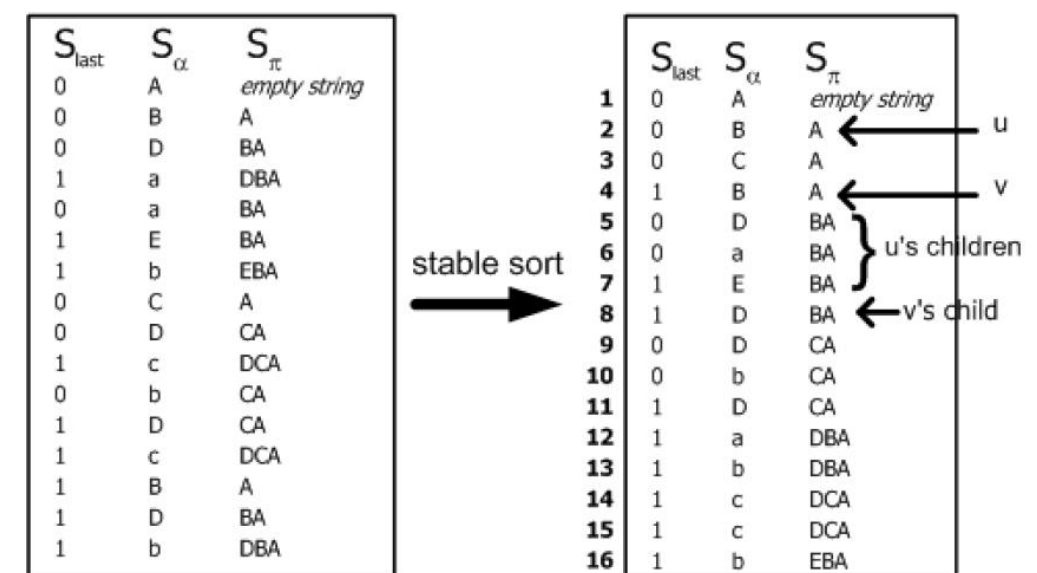
```
<biblio>
  <book id=1>
    <author>J. Austin</author>
    <title>Emma</title>
  </book>
  <book id=2>
    <author>C. Bronte</author>
    <title>Jane Eyre</title>
  </book>
</biblio>
```



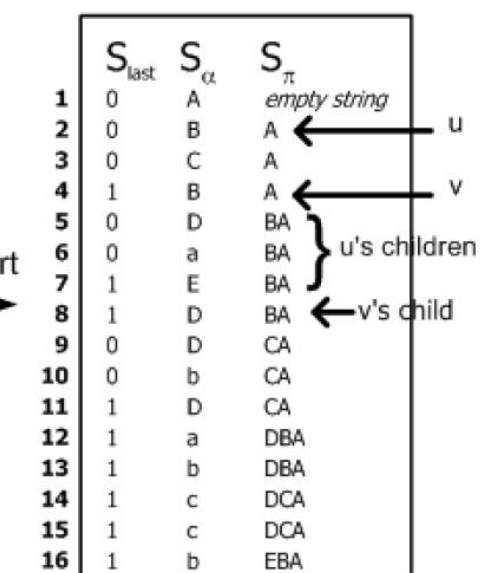
Un esempio di applicazione della tecnica brevettata



(a)



(b)



(c)

Applicabilità Industriale



DATASET	GZIP	BZIP2	PPMDI	XMILL	SCMPPM	XBZIP
PATHWAYS	7,78	4,70	3,93	10,07	3,13	1,84
DBLP	17,90	11,94	10,53	10,79	8,67	9,69
SWISSPROT	11,97	7,60	6,73	5,68	5,21	4,66
NEWS	22,94	15,06	12,62	11,54	10,71	10,71

Table 2: Comparison of XML compressors: XMILL, SCMPPM and XBZIP use PPMDI as their base compressor.

DATASET	HUFFWORD	XPRESS	XQZIP	XBZIPINDEX	XBZIP
PATHWAYS	33.68	–	–	3.62	1.84
DBLP	44.00	48	30	14.13	9.69
SWISSPROT	43.10	42	38	7.87	4.66
NEWS	45.15	–	–	13.52	10.61

Table 3: Compression ratio achieved by XML-queryable compressors over the files in our dataset. For XPRESS and XQZIP we report results taken from [39, 11] (the symbol – indicates a result not available in these papers). Note that we can trade space usage for query time by tuning the parameters of the FM-index [18].

Possibili Evoluzioni



Siamo disponibili a:

- perfezionare gli esperimenti su dataset XML e JSON di interesse per l'azienda,
- trovare altre possibili applicazioni della tecnica brevettata che considerino alberi etichettati provenienti non soltanto da XML e JSON.
- applicare la tecnica brevettata per comprimere ed effettuare ricerche su altri tipi di dati come: sequenze di byte, interi, grafi etichettati, o strutture più complesse.
- arrivare a una ingegnerizzazione software dei prototipi realizzati per ottenere i risultati sperimentali pubblicati.

Per maggiori informazioni:



Ufficio di Trasferimento Tecnologico dell'Università di Pisa

Sede: Lungarno Pacinotti 43/44, Pisa (PI) 56126

Sito web: www.unipi.it/index.php/trasferimento

E-mail: valorizzazionericerca@unipi.it

Per maggiori informazioni:



Ufficio Regionale di Trasferimento Tecnologico

Sede: Via Luigi Carlo Farini, 8 50121 Firenze (FI)

E-mail: urtt@regione.toscana.it

