



Web Mining

IL CASO DI UN NEGOZIO ELETTRONICO

- n Per valutare le abitudini sulla navigazione di un sito internet da parte degli utenti e per migliorare i link, interfaccia del sito, ottimizzando la navigazione, ecc, è possibile trovare le associazioni tra pagine visitate, cercando di agevolare ad esempio il percorso maggiormente utilizzato.
- n Un'analisi di questo tipo è detta ***click stream analysis***.

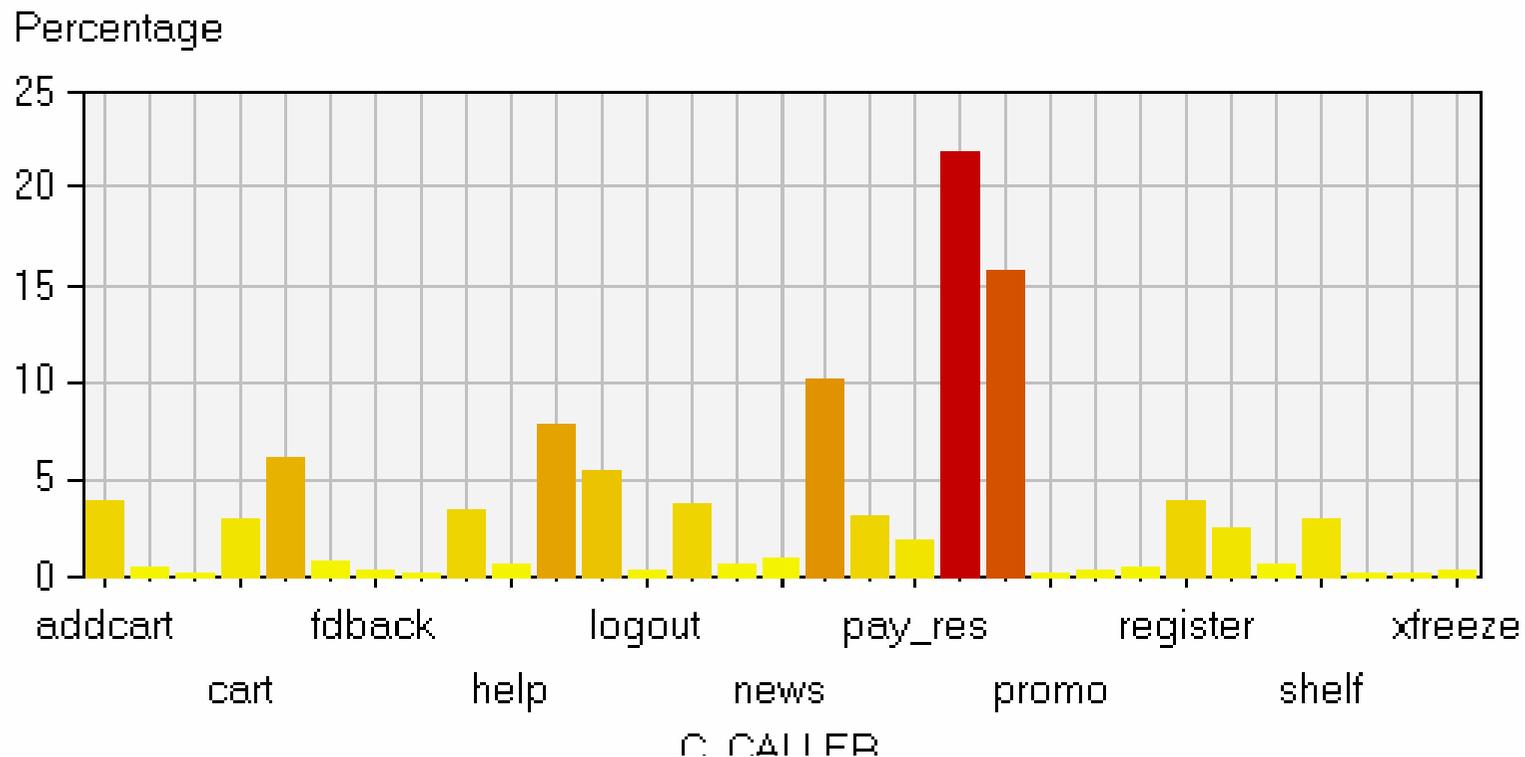
IL CASO DI UN NEGOZIO ELETTRONICO

E' stato valutato un dataset con 250.711 record, dove per ciascuno è riportato:

- q l'identificativo di ciascun utente;
 - q una variabile che indica quando è avvenuta la connessione, ossia riporta data e durata della connessione;
 - q la pagina visitata dall'utente.
-

Uso delle associazioni

Se guardiamo la distribuzione di frequenza possiamo vedere che alcune pagine risultano molto più frequentate di altre. Si desidera vedere le associazioni per facilitare il percorso e quindi l'acquisto di eventuali prodotti.



Uso delle associazioni

- n E' possibile individuare delle **catene di navigazione delle pagine, tenendo conto dell'ordine temporale** in cui sono state visitate.
- n Le associazioni che si possono avere in questo esempio sono di 1.489 con al massimo una sequenza di 3 pagine, poiché sono stati impostati un numero massimo di items nella catena più lunga pari a tre e il massimo numero di items in un'associazione.

Uso delle associazioni

	Chain Length	Support(%)	Confidence(%)	Transaction Count	Rule
248	2	2.08	4.64	469	catalog ==> regpost
249	2	2.03	10.81	458	cart ==> help
250	2	2.01	10.67	452	cart ==> agb
251	2	2.00	4.12	450	p_info ==> download
252	3	43.93	63.78	9896	program ==> product ==> product
253	3	38.46	55.84	8665	program ==> product ==> p_info
254	3	36.27	75.26	8170	product ==> p_info ==> product
255	3	33.49	84.65	7544	catalog ==> program ==> product
256	3	33.09	59.75	7455	product ==> product ==> product
257	3	30.14	84.64	6789	home ==> program ==> product
258	3	28.80	52.00	6487	product ==> product ==> p_info
259	3	28.74	75.49	6475	program ==> program ==> product
260	3	28.73	74.57	6473	program ==> p_info ==> product

Visualizzazione delle associazioni

Con la Chain Event è possibile costruire la catena di navigazione, o meglio se questa esiste, trovare la relativa confidence e support:

- q Infatti scegliendo ad esempio *catalog à program* e *programà home* si può trovare la Find Chain:
- q ***Catalog à programà home*** che ha una support del 7,64% e una confidence del 19,32%.