… qui vi raccontiamo Issue crawler che cos'è, come funziona e perché è interessante …

Presentazione di Issue Crawler

(a cura di Matteo Cernison & Claudia Padovani, Università di Padova – maggio 2006)[1]

Issue Crawler (http://www.issuecrawler.net) è un software creato dalla fondazione olandese Govcom.org e dal suo direttore Richard Rogers, docente di Nuovi Media all'Università di Amsterdam ed "epistemologo del Web", ossia ricercatore impegnato nella descrizione del Web come architettura che consente una produzione e circolazione del sapere differente rispetto a quella dei media precedenti[2]. In questa prospettiva, Rogers e i suoi colleghi si dedicano alla creazione di *info-tools,* nuovi strumenti di ricerca e analisi che offrono possibilità di indagini centrate sul Web e sulle interazioni che nello spazio da questo costituito si sviluppano.

Issue Crawler è il più famoso ed utilizzato di questi strumenti informatici. Lo scopo del programma è quello di ricostruire e presentare in forma grafica, con mappe composte da nodi interrelati, degli ***Issue Network***, reti tematiche che possono essere ricostruite fra siti (o pagine web) collegati fra loro. Attraverso l'analisi della presenza sul Web e la ricostruzione dei rispettivi rapporti (link) fra associazioni, individui, gruppi informali o istituzioni è possibile scoprire chi si occupa, ad esempio, della tutela dei diritti umani nel Sahara Occidentale o com'è strutturato l'intreccio fra i siti dei media, dei movimenti e delle realtà istituzionali che seguono il *Ciagate*. E' possibile anche, partendo da alcuni "siti di partenza", che per esempio riguardano le "mie" collaborazioni come associazione, ricostruire la "mia" rete tematica sul Web. Questo significa poter indagare quali sono gli attori che potenzialmente appartengono ad una rete tematica e quali siano la forma e le caratteristiche di ciascuna rete (centrata o meno, densa di interazioni o frammentaria…). Per rintracciare queste reti, *Issue Crawler* effettua una ***co-link analysis***, ovvero una ricerca dei collegamenti in uscita (*outgoing links*) comuni a due o più pagine web. La

[1] Presentiamo qui una descrizione alquanto semplificata, in lingua italiana, del funzionamento di Issue crawler e delle potenzialità di applicazione di questo strumento. Per ulteriori approfondimenti e spiegazioni specifiche relative alla logica di funzionamento, ai diversi interventi soggettivi implicati nell'applicazione, alla possibilità di integrare issue crawler con altri strumenti e alla lettura e analisi delle mappe fornite, rimandiamo al sito www.issuecrawler.net.

[2] Particolarmente interessante il testo di Rogers dal titolo *Information Politics on the Web*, MIT Press 2004. nel volume sono presentate anche diverse modalità di impiego degli strumenti sviluppati da Govcom.org.

co-link analysis è nata originariamente per riconoscere e ricostruire le "reti" fra pubblicazioni accademiche: al posto dei collegamenti web, venivano utilizzate le citazioni presenti nei testi, ed una pubblicazione "co-citata" da due opere appartenenti alla stessa area di studio veniva giudicata come un punto di riferimento rilevante in settore. I creatori di *Issue Crawler* sono partiti dal presupposto che anche i links, così come le citazioni accademiche, sono dei rinvii forti in grado di segnalare attori, documenti e siti riconosciuti come importanti dai creatori dei siti presi come punti di partenza. Se questi punti di partenza sono accomunati da una specifica *issue,* è molto probabile che i siti trovati attraverso la *co-link analysis* siano anch'essi legati alla stessa *issue*.

IL FUNZIONAMENTO DEL PROGRAMMA

L'algoritmo di *Issue Crawler* è piuttosto semplice. Il programma, che lavora su due server ad Amsterdam (in fase di potenziamento a causa della crescita esponenziale degli utenti negli scorsi mesi), riceve dall'utente come *input*, attraverso un'interfaccia presente nel sito www.issuecrawler.net, una serie di due o più URL di partenza, gli ***starting points*** dai quali prende il via l'analisi.

Questi punti di partenza - ad esempio siti di associazioni, istituzioni, media che trattano la *issue* su cui si vuole indagare - possono essere ricavati dalle conoscenze dall'utente, dall'interrogazione di un motore di ricerca, da pubblicazioni o interviste fatte ad esperti. Gli *starting points* possono essere anche siti di associazioni presenti contemporaneamente ad un evento[3]. Una volta lanciato il crawl, *Issue Crawler* rileva tutti i link esterni in uscita dagli *starting point* (operazione di ***Web crawling***), confronta i risultati ottenuti per ogni punto di partenza e mantiene in memoria solo i link comuni ad almeno due *starting point* (*co-link analysis*), che costituiscono l'output parziale del programma. Questo passaggio, chiamato ***iterazione***, può su indicazione dell'utente essere ripetuto una, due o tre volte. Se si sceglie di effettuare più iterazioni, la seconda e la terza utilizzano gli ultimi *co-link* memorizzati come se fossero nuovi *starting points*.

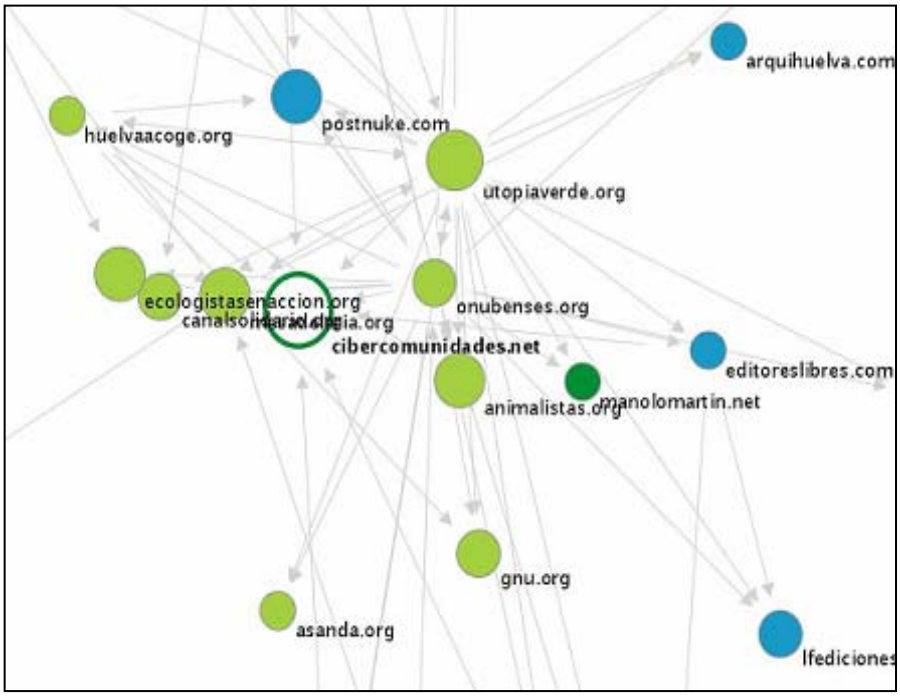
I risultati finali di questa procedura, ossia gli URLs dei co-link ottenuti nell'ultima iterazione, vengono salvati in un **file in formato XML**, archiviato nelle sezioni *Network*

[3] La scelta degli starting points è una delle operazioni più delicate, in quanto da questa scelta iniziale derivano conseguenze che possono indirizzare l'analisi in direzioni diverse. Oltre all'accuratezza nell'individuare starting point effettivamente rilevanti, è anche importante decidere da quali pagine partire per effettuare il crawl: generalmente si utilizzano non le home pages, ma le pagine dei link o delle risorse oppure pagine tematiche rilevanti per l'indagine. Questa scelta dipende, comunque, dalle domande di ricerca.

Manager e *Archive* del sito del programma. I dati contenuti in questo file (URLs dei co-link, numero di collegamenti ricevuti, per ogni co-link, dalla popolazione di siti analizzata, legami fra i vari co-link) sono sufficienti a dar vita alla mappa della rete fra i co-link, che viene generata su richiesta dell'utente e sono estremamente importanti nel condurre l'analisi sulle mappe generate dal programma.

VISUALIZZARE LE MAPPE

La mappa può essere visualizzata in due forme, come *circle map* o come *cluster map*. Nel primo caso i nodi vengono disposti in cerchio, o in due cerchi concentrici rappresentanti la periferia e il nucleo della rete, mentre nel caso delle *cluster maps* i nodi vengono posizionati in modo che la loro distanza sia inversamente proporzionale alla quantità dei loro reciproci collegamenti. Il risultato, di conseguenza, è una rete dove i siti appaiono riuniti in agglomerati, o *clusters*, facilmente individuabili sulla mappa e composti da nodi collegati da molti link reciproci.



In entrambe le forme della mappa (interattiva se salvata in formato svg), gli indirizzi Web troncati dei co-link (da http://www.euronews.com/insert1/comeri.htm si passa a euronews.com, e viene ignorato quindi il riferimento alla pagina interna a un sito) compaiono a lato dei nodi, la cui dimensione indica il "grado in entrata", ossia il numero dei collegamenti ricevuto dagli altri punti della rete. Il colore dei nodi è correlato invece al dominio di appartenenza: i siti *.com, .it, .gov* o *.org* vengono distinti così facilmente nella mappa.

Nel menù che viene visualizzato sul lato destro dello schermo sono riassunte, nella parte superiore, le principali informazioni sulla mappa e nella parte inferiore i dati

riguardanti il nodo selezionato in quel momento, fra i quali l'elenco dei co-link da cui riceve e a cui invia collegamenti.

Nel settore centrale del menù, infine, si trova la legenda interattiva dei domini presenti nella mappa e l'opzione che permette di visualizzare solamente i link che riguardano il nodo selezionato in quel momento.

E' stata da pochissimo attivata anche una terza modalità di visualizzazione: una visualizzazione sovrapposta ad una carta geografica che consente di "collocare" spazialmente i nodi che compongono la rete. Questo risulta particolarmente interessante per analisi su reti tematiche di tipo transnazionale e o regionale.

L'immagine che Issue Crawler restituisce in queste mappe è una sorta di "colpo d'occhio" sulla situazione della *issue* studiata: fotografa infatti come si presenta la *issue* in un determinato momento nello spazio estremamente dinamico del Web. Per rispondere alla mobilità della rete, e rilevare come cambia una *issue* nel tempo, Issue Crawler permette di rendere parzialmente dinamica l'osservazione e l'analisi con lo strumento *scheduler*. Con esso si può dare ordine al programma di ripetere l'analisi, partendo dagli stessi *starting point* o dagli ultimi co-link disponibili, a intervalli di tempo programmati, al fine di verificare se cambiano gli attori e la loro posizione strutturale nella rete, se si creano nuovi legami o se chi si occupa di alcune sotto-tematiche all'interno dell'issue principale cresce improvvisamente di importanza.

Alcuni links

http://www.issuecrawler.net
Interfaccia web di Issue Crawler: da qui può essere lanciato il programma, possono essere visualizzate le mappe e si accede all'insieme dei propri progetti (Sezione *Network Mamager*) e di quelli altrui (Sezione *Archive*). Nella sezione di apertura *The Lobby*, oltre alle notizie sul programma e il suo funzionamento, si trovano informazioni sull'andamento e la durata prevista dei *crawls* in corso, e si può controllare la posizione dei propri crawls nella coda di attesa.

Da http://www.govcom.org, sito della fondazione diretta da Richard Rogers, si può accedere fra l'altro a:

→ http://www.govcom.org/Issuecrawler_instructions.htm
Istruzioni di Issue Crawler
→http://www.govcom.org/publications.html
→http://www.govcom.org/full_list.html
→http://www.govcom.org/drafts.html
Articoli, mappe e ricerche effettuate dalla Govcom.org utilizzando i propri strumenti web, e principalmente Issue Crawler
→http://www.govcom.org/Issue Crawler Scenarios of Use.htm

Scenari ed indicazioni su come adattare il programma e i suoi settaggi alla propria ricerca

L'articolo all'URL http://www.easst.net/rewiev/sept2002/issuecrawler è una presentazione di Richard Rogers al programma e al progetto. Altre informazioni si trovano nella pagina "wiki" http://wiki.issuecrawler.net/bin/view/Issuecrawler/WebHome.

http://movies.issuecrawler.net ospita un filmato dove il funzionamento di Issue Crawler viene spiegato mostrando come esempio una ricerca sui movimenti sociali legati al problema della costruzione di dighe nella valle del fiume Narmada, in India.

In http://tools.issuecrawler.net sono presenti, infine, alcuni strumenti affiancati spesso nelle ricerche a Issue Crawler.