

PageRank 簡介

([臺大圖書館參考服務部落格](#) [《SJR—SCOPUS 期刊評鑑指標\(1\)》](#))

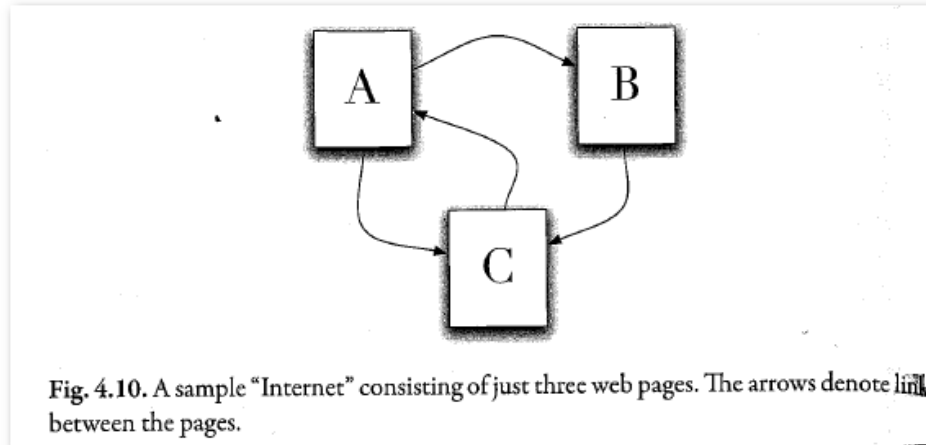
與 [《Eigenfactor 與 Article Influence 介紹—JCR 期刊評鑑指標\(2\)》](#) 延伸閱讀)

搜尋引擎要從成先上萬的網頁中篩選出受歡迎者，最簡單的方法是計算各網頁的被連結數(inlinks)，因為超連結可以視為對被連結網站的信任投票，因此得票數越高的網站可以推估其內容越重要，值得被排序在較前面。

然而，被連結數易被操控，例如網站經營者可能為了提高自己的能見度而創造大量垃圾連結指向同一目標網站，藉此提高被連結數。另外，單純計算被連結數，無法有效給予每個連結相對的權重，例如被獲得知名網站的連結與獲得一般網站的連結，其重要性應該要有所區別。基於上述理由，在考量連結因素方面更可靠的評估方法如 PageRank 便因運而生。

PageRank 是指網頁被看到的可能性，每個網頁都有個別的 PageRank，取決於網頁間連結關係，受歡迎的網頁藉此脫穎而出。其概念源自於網路上的「隨機瀏覽」(random surfer)行為。假使一個網路使用者不斷的瀏覽網站，其瀏覽方式為透過目前瀏覽的頁面之超連結，到達下一個網站。當他目前瀏覽的網站中沒有任何超連結、或對其超連結內容都不感興趣時，使用者會跳到任一隨機的頁面去繼續瀏覽。

在這樣隨機瀏覽的過程中，受歡迎的網頁容易被看到，因為大多數的網頁傾向連結受歡迎的網頁；被受歡迎網頁連結的網頁，能見度也大幅提升。為了要達到此種重要性、連結結構、與能見度之間交互影響的結果，PageRank 的運算公式被設計為「一個網站的 PageRank 值，來自於加總所有連結到該網站的網站之 PageRank 值除以本身的導出連結數」。以下述情形為例，假設網路上僅有 A、B、C 三個網頁，其互相連結的關係如下圖：



則 C 網站的 PageRank 值來自於連結到 C 的網站，也就是網站 A 和 B，將 A 和 B 的 PageRank 值除以其各自的導出連結數，再將此二數值加總，即可得 C 網站的 PageRank 值如下：

$$PR(C) = \frac{PR(A)}{2} + \frac{PR(B)}{1}$$

若將此公式其一般化，則可得 PageRank 值的計算方式為：

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L_v}$$

此公式是一個會收斂的運算。以上述例子而言，一開始假設每個網頁的 PageRank 值都是均等的，則計算方法如下(每階段的 PR 值使用前一階段的運算結果)：

(1) $PR(A)=PR(B)=PR(C)=1/3=0.33$

(2) $PR(A)=0.33 \quad PR(B)=0.33/2=0.17 \quad PR(C)=0.33/2+0.33=0.5$

(註：B 獲得 A 的連結，而 A 有 2 個導出連結，因此 $PR(B)=PR(A)/2$ ；C 獲得 A 和 B 的連結，A 有 2 個導出連結，B 只有 1 個，因此 $PR(C)=PR(A)/2+PR(B)$)

$$(3) \quad PR(A)=0.5 \quad PR(B)=0.33/2=0.17 \quad PR(C)=0.33/2+0.17=0.33$$

(註：A 獲得 C 的連結，C 只有 1 個導出連結，因此 $PR(A)$ =前一階段 $PR(C)$ 的運算結果；B 獲得 A 的連結，而 A 有 2 個導出連結，因此 $PR(B)$ =前一階段的 $PR(A)/2$ ；C 獲得 A 和 B 的連結，A 有兩個導出連結，B 只有一個，因此 $PR(C)$ =前一階段 $PR(A)/2$ +前一階段 $PR(B)$)

$$(4) \quad PR(A)=0.33 \quad PR(B)=0.5/2=0.25 \quad PR(C)=0.5/2+0.17=0.42$$

(5) 依此類推...

最後趨近：

$$PR(A)=0.4 \quad PR(B)=0.2 \quad PR(C)=0.4$$

若每個網頁被瀏覽到的機率，再進一步加入不是透過其它網頁的連結而來，而是透過某個隨機連結而來，假設透過隨機連結而來的機率為 λ ，則 C 網站的 PageRank 值計算，更換如下：

$$PR(C) = \frac{\lambda}{3} + (1 - \lambda) \cdot \left(\frac{PR(A)}{2} + \frac{PR(B)}{1} \right)$$

若將此公式其一般化，則可得 PageRank 值的計算方式為：

$$PR(u) = \frac{\lambda}{N} + (1 - \lambda) \cdot \sum_{v \in B_u} \frac{PR(v)}{L_v}$$

此公式落實了 PageRank 值的發想，意即每個網頁的 PageRank 值與連結到該網頁的網頁之 PageRank 值成正相關，因此被越多網頁連結到可以獲得越多的 PageRank 值累加，且被 PageRank 值越高的網站連結可獲得越大的 PageRank 值挹注。而一個網頁若有越少的導出連結，代表來到該網頁的使用者進入任一其連

結的網頁的機率越大，因此其能挹注給個別導出連結網頁的 PageRank 值也將越大。而由上述的 PageRank 理論公式，每個網頁經計算後可得到一個介於 0~1 間的值，然而這與我們一般所熟知 Google 介於 1~10 等級的 PageRank 值有所不同，搜尋引擎將透過後端的計算將所得的 PageRank 值轉換為 1~10 的等級呈現。

資料來源：Croft, Bruce; Metzler, Donald; Strohman, Trevor, Search Engines: Information Retrieval in Practice, Addison Wesley, 2010.[[臺大圖書館館藏](#)]

※返回閱讀《[SJR—SCOPUS 期刊評鑑指標\(1\)](#)》

與《[Eigenfactor 與 Article Influence 介紹—JCR 期刊評鑑指標\(2\)](#)》※