

**Δημοσιευμένο στο :**

Μιχάλης Τζεκάκης : ο οραματιστής. Αθήνα : Ένωση Ελλήνων Βιβλιοθηκονόμων και Επιστημόνων της Πληροφόρησης, Xerox, 2007. σ. 44-63

## **Βιβλιομετρική-Πληροφοριομετρική-Ιστομετρική**

**Ανέστης Σίτας**

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης ([sitas@lit.auth.gr](mailto:sitas@lit.auth.gr))

**Ουρανία Κώνστα**

Ιόνιο Πανεπιστήμιο ([rania@ionio.gr](mailto:rania@ionio.gr))

**Περσεφόνη Ντούλια**

Πανεπιστήμιο Ιωαννίνων ([pntoylia@yahoo.gr](mailto:pntoylia@yahoo.gr))

## **Bibliometrics-Informetrics-Webometrics**

### **Abstract**

The information available on the Web has been highly increased. At the same time Web becomes the main medium of publication and distribution of information. This creates a need for exploration, study, interpretation and presentation of its characteristics. Simultaneously a need is created for the study of patterns of behaviors, practices, models and tools of Web users. The results of surveys lead us to the understanding, improvement and the appropriate exploitation of the Web for the scientific research. Methods, techniques and tools that are used originated in the application of bibliometric studies in traditional scientific publications, have reached the new and explosively growing field of Webometrics. Webometrics concerns the application of quantitative methods for the exploration of web pages and web sites.

### **Περίληψη**

Οι διαθέσιμες πληροφορίες στο Web αυξάνονται γεωμετρικά. Συγχρόνως το Web γίνεται το κύριο μέσο δημοσίευσης και διακίνησης πληροφοριών. Αυτό δημιουργεί την ανάγκη εξερεύνησης, μελέτης, ερμηνείας και παρουσίασης των χαρακτηριστικών του. Ταυτόχρονα δημιουργείται η ανάγκη για μελέτη καθορισμένων συμπεριφορών, πρακτικών, μοντέλων και εργαλείων που εφαρμόζουν οι χρήστες του. Τα αποτελέσματα των ερευνών αυτών οδηγούν στην κατανόηση του μέσου, καθώς και στη βελτίωση και την κατάλληλη αξιοποίησή του για την επιστημονική έρευνα. Οι μέθοδοι, οι τεχνικές και τα εργαλεία που χρησιμοποιούνται ξεκίνησαν από την εφαρμογή βιβλιομετρικών ερευνών σε παραδοσιακές επιστημονικές δημοσιεύσεις και έφθασαν ως το νέο και διαρκώς ευρύτερο πεδίο της Webometrics, που αφορά την εξερεύνηση των τεκμηρίων, ιστοσελίδων και δικτυακών τόπων του Web.

### **1.1. Βιβλιομετρική**

Για πολλά χρόνια η βιβλιομετρική, η εφαρμογή δηλαδή των ποσοτικών τεχνικών έρευνας και της στατιστικής στα δημοσιεύματα ενός τομέα ή συνόλου βιβλιογραφίας, υπήρξε μια σημαντική περιοχή μελέτης των επιστημών της πληροφόρησης. Είναι το επιστημονικό πεδίο που αναπτύχθηκε κατά το πρώτο μισό του 20ου αιώνα, από το ενδιαφέρον των ερευνητών για τη δυναμική της επιστήμης

όπως αντανακλάται στη βιβλιογραφική παραγωγή της. Το ενδιαφέρον αυτό σχετίζεται με την καταγραφή αλλαγών στα επιστημονικά πεδία μέσα στο χρόνο ή στις διάφορες χώρες, με το πρόβλημα που αντιμετωπίζουν οι βιβλιοθήκες στο χειρισμό αυτών των αποτελεσμάτων και με τη χαμηλή παραγωγή δημοσιευμάτων πολλών επιστημόνων. Είναι ένα είδος ερευνητικής μεθόδου που χρησιμοποιείται στην επιστήμη της πληροφόρησης.

Η εξέλιξη της βιβλιομετρικής χαρακτηρίστηκε από δύο βασικά στοιχεία: α) πως οι (επιστημονικές) θεματικές βιβλιογραφίες, δηλαδή οι δευτερογενείς βιβλιογραφικές πηγές, περιέχουν όλα τα στοιχεία της βιβλιογραφίας που είναι απαραίτητα για την ποσοτική ανάλυση και παράλληλα εξασφαλίζουν οικονομία χρόνου, και β) ο προσανατολισμός των επιστημόνων στην αναζήτηση αριθμητικών στοιχείων και η ανάλυσή τους με στατιστικές ή άλλες μαθηματικές μεθόδους.

Αρχικά οι προσπάθειες αυτές αναφερόταν ως «στατιστική βιβλιογραφία». Το 1969 προτάθηκε από τον A. Pritchard ως πιο κατάλληλος ο όρος «βιβλιομετρική», που την ορίζει ως «την εφαρμογή μαθηματικών και στατιστικών μεθόδων στα βιβλία και σε άλλα μέσα επικοινωνίας». Την ίδια χρονιά από τον R. A. Fairthorne ορίζεται ως «ο ποσοτικός χειρισμός των ιδιοτήτων του καταγεγραμμένου συνεχούς λόγου (γραπτού και προφορικού) και των εμπεριεχομένων συμπεριφορών». Το 1989 από τους Hume και MacCain η βιβλιομετρική ορίζεται ως «η ποσοτική μελέτη των βιβλιογραφικών παραπομπών όπως αυτές εμφανίζονται στις βιβλιογραφίες, με στόχο την παροχή εξελικτικών μοντέλων στις επιστήμες και την τεχνολογία. (Wilson, 1999: 108).

Η βιβλιομετρική εμπεριέχει δύο συστατικά. Το πρώτο συστατικό της είναι η αναζήτηση συστηματικά επαναλαμβανόμενων μοντέλων (patterns) σε ολοκληρωμένα και ομοειδή σώματα βιβλιογραφίας. Το δεύτερο είναι καταγραφή ειδικών πεδίων ή θεμάτων που χρησιμοποιούν αυτό το πλαίσιο των δεικτών και νόμων. Ένας κοινός τρόπος για τη βιβλιομετρική έρευνα είναι η χρησιμοποίηση του Science Citation Index, του Social Science Citation Index ή του Arts and Humanities Citation Index για τον εντοπισμό και την μελέτη των παραπομπών.

## **1. 2. Ανάλυση βιβλιογραφικών παραπομπών (citation analysis)**

Η ανάλυση των βιβλιογραφικών παραπομπών επηρέασε σε μεγάλο βαθμό το σύγχρονο πεδίο της ανάκτησης πληροφοριών. Από κάποιους ερευνητές παρουσιάζεται ως η «μεγάλη αδελφή» της βιβλιομετρικής, ενώ από άλλους ως ένα από τα κύρια συστατικά της (Wilson, 1999: 108). Ως λόγοι μελέτης των παραπομπών μπορεί να αναφερθούν οι παρακάτω (Turnbull, 2007):

- Η απόδοση φόρου τιμής στους πρωτοπόρους ενός επιστημονικού κλάδου
- Ο εντοπισμός σχετιζόμενων έργων
- Ο εντοπισμός μεθοδολογιών, εργαλείων, κτλ.
- Ο εντοπισμός σχέσεων μεταξύ των συγγραφέων και των έργων τους
- Η κριτική δημοσιευμένων έργων
- Η τεκμηρίωση απόψεων
- Η παροχή ώθησης σε έργα που δεν είναι πολύ γνωστά, ευρητηριασμένα ή δεν έχουν χρησιμοποιηθεί ως παραπομπές
- Η δημιουργία στοιχείων επικύρωσης των δεδομένων, των γεγονότων και των αποτελεσμάτων ερευνών
- Ο προσδιορισμός των πρωτογενών δημοσιεύσεων στα οποία περιέχονται νέες ιδέες και στοιχεία
- Ο προσδιορισμός των πρωτογενών δημοσιεύσεων ή άλλων έργων που περιγράφουν τα χαρακτηριστικά γνωρίσματα μιας ιδέας ή ενός όρου

- Η αντίκρουση ή αμφισβήτηση έργων ή ιδεών διαφόρων συγγραφέων (αρνητικές απόψεις)

Η μελέτη ζευγών παραπομπών (co-citation coupling) και η βιβλιογραφική ζεύξη (Bibliographic coupling) αποτελούν είδη του βασικού πεδίου της ανάλυσης βιβλιογραφικών παραπομπών.

- *Ανάλυση παραπομπών.* Όταν ένας συγγραφέας παραπέμπει σε έναν άλλο, καθιερώνεται μια σχέση. Η ανάλυση των βιβλιογραφικών παραπομπών χρησιμοποιεί τις παραπομπές των έργων για να δημιουργήσει συνδέσεις/συσχετίσεις. Μπορούν να εξακριβωθούν πολλές διαφορετικές συσχετίσεις, όπως μεταξύ συγγραφέων, έργων, περιοδικών, τομέων ή ακόμα και μεταξύ χωρών. Μπορούν να μελετηθούν και οι παραπομπές από και σε ένα ορισμένο τεκμήριο. Μία συνηθισμένη χρήση της ανάλυσης παραπομπών είναι για τον προσδιορισμό της επιρροής που έχει ασκήσει ένας συγγραφέας σε ένα τομέα, υπολογίζοντας τις αναφορές που έχουν γίνει σε αυτόν από άλλους συγγραφείς. Ένα πιθανό μειονέκτημα αυτής της προσέγγισης είναι ότι οι συγγραφείς μπορεί να αναφέρουν τον συγκεκριμένο συγγραφέα σε ένα αρνητικό πλαίσιο (π.χ. να λένε ότι ο συγγραφέας δεν γνωρίζει το θέμα).
- *Co-citation coupling (ζεύγη παραπομπών).* Είναι μία μέθοδος που χρησιμοποιείται για να καθιερώσει μία θεματική ομοιότητα μεταξύ δύο τεκμηρίων. Εάν οι δημοσιεύσεις Α και Β αναφέρονται στη δημοσίευση Γ, θεωρούνται σχετικές μεταξύ τους, έστω κι αν δεν συνδέονται άμεσα. Αν οι δημοσιεύσεις Α και Β αναφέρονται μαζί σε πολλές άλλες, τότε έχουν μία ισχυρότερη σχέση. Σε όσες περισσότερες δημοσιεύσεις αναφέρονται, τόσο ισχυρότερη είναι η σχέση τους.
- *Bibliographic coupling (βιβλιογραφική ζεύξη).* Λειτουργεί με μία παρόμοια αρχή, αλλά αποτελεί τον αντικατοπτρισμό των ζευγών παραπομπών. Συνδέει δύο δημοσιεύσεις που αναφέρουν τα ίδια άρθρα, έτσι ώστε εάν οι δημοσιεύσεις Α και Β αναφέρουν τη δημοσίευση Γ, τότε μπορούμε να πούμε ότι είναι σχετικές, έστω και αν δεν συνδέονται άμεσα μεταξύ τους. Όσα περισσότερες κοινές δημοσιεύσεις αναφέρουν τόσο ισχυρότερη είναι η σχέση τους.

### 1. 3. Επιστημονομετρική (Scientometrics)

Ο όρος αυτός προτάθηκε το 1969 από τους Nalimov και Mulchenko. Ορίζεται ως «η ποσοτική μελέτη της επιστήμης και της τεχνολογίας». Η επιστημονομετρική δύσκολα ξεχωρίζει από τη βιβλιομετρική. Η βιβλιομετρική εστιάζει κυρίως το ενδιαφέρον της σ' αυτή καθ' αυτή την βιβλιογραφία των διαφόρων επιστημών, ενώ την επιστημονομετρική την απασχολούν πολλές και διαφορετικές μετρήσεις και αναλύσεις στο πεδίο των επιστημών και της τεχνολογίας πέρα από τη βιβλιογραφία (π.χ. οι πρακτικές των επιστημόνων, οι κοινωνικο-οργανωτικές δομές, η διοίκηση έρευνας και ανάπτυξης, κτλ) (Wilson, 1999:108).

### 1. 4. Μελέτες επιστημονικής επικοινωνίας (Scholarly communication studies)

Εκτός από την επιστημονομετρική υπάρχουν και άλλα είδη μελετών που επικαλύπτουν την βιβλιομετρική, όπως για παράδειγμα η μελέτες επιστημονικής επικοινωνίας. Το πεδίο αυτό αφορά τη μελέτη του τρόπου που οι επιστήμονες διαφόρων πεδίων χρησιμοποιούν και διαχέουν τις πληροφορίες μέσω επισήμων ή ανεπισήμων καναλιών. Στην περίπτωση αυτή η βιβλιομετρικές μέθοδοι εφαρμόζονται μόνο στην περίπτωση των επισήμων καναλιών.

### 1. 5. Πληροφοριομετρική (Informetrics)

Ο όρος αυτός προτάθηκε το 1979 από τον O. Nacke για να ορίζει το τμήμα της επιστήμης των πληροφοριών που ασχολείται με τις μετρήσεις των πληροφοριακών φαινομένων και την εφαρμογή μαθηματικών μοντέλων σ' αυτά. Το 1984 από την FID (Federation Internationale de Documentation) χρησιμοποιείται ως κοινός όρος για τη βιβλιομετρική και την επιστημονομετρική. Το 1988 από τους Egghe, Rousseau και Brooks ορίζεται ως γενικός όρος για την επιστημονομετρική και τη βιβλιομετρική, με την πρώτη να δίνει έμφαση στις μελέτες που ασχολούνται με την ακολουθούμενη πολιτική, και τη δεύτερη με μεγαλύτερη έμφαση στην επιστήμη της βιβλιοθηκονομίας (Wilson, 1999: 112). Η πληροφοριομετρική αφορά τη μελέτη των ποσοτικών πλευρών της πληροφορίας σε κάθε της μορφή και σε κάθε κοινωνική ομάδα. Ερευνά τα χαρακτηριστικά και τις μετρήσεις προσώπων, ομάδων, ιδρυμάτων, χωρών, τα δημοσιεύματα και τις πληροφοριακές πηγές, τις θεματικές περιοχές και τις διαδικασίες ανάκτησης.

Όταν δόθηκαν οι παραπάνω ορισμοί δεν υπήρχε ακόμη το Web, σήμερα όμως που όλο και σε μεγαλύτερο βαθμό αποτελεί την κύρια πηγή πληροφόρησης, οι μέθοδοι αυτοί μπορούν να εφαρμοστούν και σ' αυτό το πεδίο και συγχρόνως αναπτύσσονται και νέες μέθοδοι εξειδικευμένες στο Web (Bar, et al, 2002).

## **1. 6. Βιβλιομετρική-πληροφοριομετρική-webometrics (ιστομετρική ;)**

Με την τεχνολογική εξέλιξη και τη διαφαινόμενη μετάβαση από το έντυπο στο ηλεκτρονικό τεκμήριο, οι βιβλιομετρικές μελέτες έδωσαν την εντύπωση πως θα αναφερόταν πλέον σε ένα διαρκώς μειούμενο πεδίο, αφού λογικά θα έπρεπε να αναφέρονται σε παραδοσιακά έντυπα τεκμήρια. Ουσιαστικά όμως παρουσιάζεται μια μετονομασία, ή αντικατάσταση σύμφωνα με άλλους ερευνητές, της «βιβλιομετρικής» σε «πληροφοριομετρική» χωρίς ουσιαστικές αλλαγές στο αντικείμενο του ενδιαφέροντός τους, ώστε να καλυφθεί η αλλαγή που επήλθε στο υλικό που δημοσιεύεται. Ίσως τελικά η «βιβλιομετρική» να πρέπει να θεωρηθεί ως ο προάγγελος της «πληροφοριομετρικής» η οποία με την σειρά της γίνεται ο προάγγελος της webometrics, που αφορά τη χρήση μεθόδων της πληροφοριομετρικής στο Web (Wilson, 1999: 108).

## **2. Νόμοι της βιβλιομετρικής**

Οι νόμοι αυτοί αυτούσιοι ή με τις κατάλληλες προσαρμογές εφαρμόζονται και στην πληροφοριομετρική και στην webometrics.

Ο **νόμος του Lotka** (ο Lotka ήταν βιομαθηματικός) περιγράφει τη συχνότητα δημοσίευσης των συγγραφέων ενός συγκεκριμένου τομέα. Δηλώνει ότι «ο αριθμός των συγγραφέων που έχει παράγει  $n$  άρθρα είναι περίπου  $1/n^2$  αυτών που δημιουργούν μόνο ένα και η αναλογία των συνεισφερόντων, που έχουν μία συμβολή, είναι περίπου 60%. Αυτό σημαίνει ότι από όλους τους συγγραφείς σε ένα συγκεκριμένο τομέα, 60% θα έχουν μόνο μία δημοσίευση και 15% θα έχουν δύο δημοσιεύσεις ( $1/2^2$  επί 60), 7% των συγγραφέων θα έχουν τρεις δημοσιεύσεις ( $1/3^2$  επί 60) και ούτω καθ' εξής. Σύμφωνα με τον νόμο της επιστημονικής παραγωγικότητας του Lotka, μόνο 6% των συγγραφέων σε ένα τομέα θα παράγουν περισσότερα από δέκα άρθρα. Όταν εφαρμόζεται σε μεγάλους όγκους βιβλιογραφίας και για μεγάλη χρονική περίοδο μπορεί να είναι γενικά αξιόπιστος, αλλά όχι στατιστικά ακριβής. Χρησιμοποιείται συχνά για να υπολογίσει τη συχνότητα με την οποία οι συγγραφείς θα εμφανιστούν σε έναν on line κατάλογο (Palmquist, 2004).

Ο νόμος του Bradford-νόμος της διασποράς (ο Bradford ήταν χημικός που αργότερα ασχολήθηκε με τη βιβλιοθηκονομία), χρησιμεύει ως μία γενική οδηγία προς τους βιβλιοθηκονόμους για τον προσδιορισμό των βασικών τίτλων περιοδικών σε μία συγκεκριμένη θεματική περιοχή. Δηλώνει ότι τα περιοδικά σε έναν συγκεκριμένο θεματικό πεδίο μπορούν να διαιρεθούν σε τρεις ομάδες, η κάθε μία εκ των οποίων περιέχει τον ίδιο αριθμό άρθρων:

- 1) η πρώτη βασική ομάδα αποτελείται από έναν πυρήνα τίτλων περιοδικών για το πεδίο αυτό, σχετικά λίγοι σε αριθμό, που αποτελεί περίπου το ένα τρίτο του συνόλου των άρθρων,
- 2) η δεύτερη ομάδα περιέχει τον ίδιο αριθμό άρθρων με την πρώτη, αλλά ένα μεγαλύτερο αριθμό τίτλων περιοδικών, και
- 3) η τρίτη ομάδα, περιέχει τον ίδιο αριθμό άρθρων με τη δεύτερη και ένα ακόμη μεγαλύτερο αριθμό τίτλων περιοδικών.

Η μαθηματική σχέση του αριθμού των περιοδικών στον πυρήνα της πρώτης ομάδας είναι μία σταθερά  $n$  και στη δεύτερη ζώνη η σχέση είναι  $n^2$ . Ο Bradford εξέφρασε πως ο αριθμός των περιοδικών στις τρεις παραπάνω ομάδες αυξάνει με την αναλογία  $1 : n : n^2$ , (π.χ. αν έχουμε μια ομάδα 3000 άρθρων που αναφέρονται πάνω σε ένα θέμα δημοσιευμένων σε 800 περιοδικά, τότε η αναλογία σε συνεχόμενες ομάδες των 1000 άρθρων, θα ήταν  $42 : 158 : 600$  ή περίπου  $1 : 3.8 : 3.8^2$ ) (Wilson, 1999: 166). Είναι φανερό ότι ο νόμος του Bradford δεν είναι στατιστικά ακριβής αλλά συνήθως χρησιμοποιείται ως γενικός κανόνας ένδειξης.

Ο νόμος του Zipf-περιστατικά λέξεων (Ο Zipf ήταν φιλόλογος) χρησιμοποιείται για να προβλέψει τη συχνότητα της εμφάνισης των λέξεων μέσα σε ένα κείμενο. Δηλώνει ότι “αν οι λέξεις ταξινομηθούν κατά φθίνουσα σειρά του αριθμού εμφάνισής τους σε ένα σχετικά μεγάλο κείμενο, τότε η θέση/σειρά μιας λέξης σε αυτόν τον κατάλογο όταν πολλαπλασιάζεται με τη συχνότητα εμφάνισής της είναι ίση με μια σταθερά. Η εξίσωση για αυτή τη σχέση είναι:  $r \times f = k$ , όπου  $r$  είναι η θέση/σειρά της λέξης,  $f$  είναι η συχνότητα και  $k$  είναι η σταθερά (Palmquist, 2004).

Για παράδειγμα αν μέσα σε ένα αρκετά μεγάλο κείμενο μια λέξη που βρίσκεται στην θέση 1 (π.χ. η λέξη the) εμφανίζεται 1000 φορές, τότε η λέξη της θέσης 2 (π.χ. η λέξη of) θα εμφανίζεται περίπου 500 φορές, η λέξη της θέσης 3 (π.χ. η λέξη and) θα εμφανίζεται περίπου 300 φορές, κτλ. (Wilson, 1999: 166). Ο νόμος του Zipf δεν είναι στατιστικά τέλειος, αλλά είναι χρήσιμος στους ευρετηριαστές.

Εκτός των παραπάνω νόμων εφαρμόζονται και άλλοι ειδικότεροι νόμοι οι οποίοι συνήθως καλύπτουν πιο εξειδικευμένες περιπτώσεις. Μερικοί από αυτούς είναι (Turnbull, 2007):

- Ο αντίστροφος νόμος της ανάκλησης και της ακρίβειας του Cleverton (Clevertons’s Inverse Law of Recall and Precision). Εφαρμόζεται για τον συνδυασμό μιας ομάδας όρων (με επεξεργασία που γίνεται με παραδοσιακό τρόπο) και τη χρήση παραδοσιακών τεχνικών για την σταδιακή αντικατάσταση της παραδοσιακής (ανθρώπινης) ευρετηρίασης με στατιστικές τεχνικές. Μπορεί να εφαρμοστεί σε περιβάλλον Web με τη χρήση σταθμισμένων πινάκων (weighted tables) καθώς και με την χρήση συστημάτων fuzzy logic.
- Η Γενική Θεωρία των Συστημάτων Πληροφόρησης του Goffman. Σύμφωνα με τη θεωρία αυτή οι ιδέες παρουσιάζονται σε συγκεκριμένους τομείς σε τακτά χρονικά διαστήματα. Το ίδιο συμβαίνει και στην έρευνα. Μπορούν δηλαδή να χρησιμοποιηθούν όλες οι παραπάνω τεχνικές για να γίνουν οι μετρήσεις που απαιτούνται από την κάθε θεωρία. Μπορεί να χρησιμοποιηθεί λογισμικό για την

μελέτη των περιλήψεων, των παραπομπών και της βιβλιογραφίας. Στη συνέχεια τα αποτελέσματα τοποθετούνται σε πίνακες που βοηθούν στην εύρεση συσχετίσεων.

### 3. Το περιεχόμενο της πληροφοριομετρικής

Η μελέτη της πληροφοριομετρικής έρευνας μπορεί να ταξινομηθεί με διάφορους τρόπους, για παράδειγμα ανάλογα με τον τύπο των δεδομένων (παραπομπές, συγγραφείς, όροι ευρετηρίασης, κτλ.), τη μέθοδο ανάλυσης των δεδομένων που χρησιμοποιείται (στατιστικά συχνότητας, ανάλυση ομάδων-cluster analysis) ή τον τύπο των επιδιωκόμενων στόχων και των αποτελεσμάτων που επιτεύχθηκαν (δείκτες αξιολόγησης, δομή ή περιγραφικά στατιστικά).

Το βασικό αντικείμενο μελέτης της είναι μια συγκεκριμένη συλλογή δημοσιευμάτων, συνήθως άρθρα, περιοδικά ή μονογραφίες, αλλά μπορεί να είναι και ολόκληρα περιοδικά, πρακτικά συνεδρίων ή βάσεις δεδομένων. Κάθε δημοσίευμα θεωρείται ως αποθετήριο ιδιοτήτων (βιβλιογραφικών πεδίων) με διάφορες τιμές, για παράδειγμα γλώσσα, έτος έκδοσης, περιοδικό δημοσίευσης, συγγραφείς και τίτλοι. Κάθε μια από τις ιδιότητες αυτές περιέχει άλλες ιδιότητες, όπως ο επιμελητής του περιοδικού στο οποίο δημοσιεύτηκε το άρθρο, το ίδρυμα από το οποίο προέρχεται ο συγγραφέας του, κτλ.

Ένας άλλος περισσότερο πολύπλοκος τρόπος, είναι η μελέτη των τεκμηρίων σε άξονες σχέσεων (π.χ. περιέχεται στο, γράφηκε από τον) με τις τιμές των ιδιοτήτων ή σχετιζόμενα αντικείμενα (quasi-objects) (π.χ. περιοδικό X, συγγραφέας Y). Το κύριο συστατικό αυτού του μοντέλου είναι ένα σύνολο από πραγματικές συνδέσεις μεταξύ των δημοσιευμάτων, που μπορούν να ξεκινούν από βιβλιογραφικές παραπομπές από δημοσιεύματα και να φτάνουν ως βιβλιογραφικές παραπομπές σε δημοσιεύματα, με το σύνολο των δημοσιευμάτων της άλλης σύνδεσης πιθανά να επεκτείνει την αρχική συλλογή. Αφού τα δημοσιεύματα στην κάθε πλευρά της σύνδεσης είναι από μόνα τους αποθετήρια ιδιοτήτων, μπορεί να γίνουν έμμεσες συγκρίσεις μέσω των συνδέσεων, π.χ. από το συγγραφέα ενός δημοσιεύματος στο συγγραφέα ενός άλλου δημοσιεύματος που έχει χρησιμοποιηθεί ως παραπομπή ή από το έτος δημοσίευσης μιας εργασίας στο έτος δημοσίευσης των μονογραφιών που χρησιμοποιούνται ως παραπομπές σ' αυτήν.

Άλλα τέσσερα στοιχεία είναι απαραίτητα να αναφερθούν (Wilson 1999: 118):

- **Πρώτο**, η βασική συλλογή δημοσιευμάτων προσδιορίζεται με βάση επιλεγμένες τιμές των ιδιοτήτων των δημοσιευμάτων αυτών, π.χ. μπορούν να μελετηθούν οι δημοσιεύσεις σχετικές με ένα θέμα ή που έχουν παραχθεί από ένα συγκεκριμένο οργανισμό μέσα σε ένα συγκεκριμένο χρονικό διάστημα
- **Δεύτερο**, κάθε μεμονωμένο δημοσίευμα μπορεί να οριστεί, έστω και τυπικά, με μια ενδεικτική ιδιότητα που συνήθως αφορά κάποιου είδους συνδυασμό στοιχείων άλλων ιδιοτήτων, όπως για παράδειγμα συγγραφέας-έτος-περιοδικό-πρώτη σελίδα ή απλώς με έναν αριθμό εισαγωγής μέσα στην συλλογή.
- **Τρίτο**, οι ιδιότητες των δημοσιευμάτων μπορεί να είναι εγγενής, με την έννοια πως σε ένα συγκεκριμένο δημοσίευμα μπορεί να καθοριστεί η τιμή του απ' ευθείας από τα στοιχεία του κειμένου του, ή να είναι εξωγενής, π.χ. δεδομένα χρήσης σε βιβλιοθήκες, περιγραφείς που δίνονται από υπηρεσίες δημιουργίας περιλήψεων και ευρετηρίασης, ή ομάδες παραπομπών όπως αυτές παρουσιάζονται σε ευρετήρια παραπομπών. Αυτή η διάκριση ίσως να είναι και αμφισβητήσιμη, αφού η πληροφοριομετρική έρευνα βασίζεται στις βιβλιογραφίες των αναγραφών και όχι απ' ευθείας στα δημοσιεύματα αυτά καθ' αυτά.
- **Τέταρτο**, πολλές ιδιότητες μπορούν να ταξινομηθούν ονομαστικά, π.χ. η γλώσσα παίρνει την τιμή Ελληνικά, Αγγλικά, Γερμανικά, κτλ.

Στην πληροφοριομετρική ανάλυση, σχετιζόμενες αριθμητικές ιδιότητες δημιουργούνται με τη σύσταση διαφόρων υποσυνόλων με κοινές ονομαστικές αξίες και στη συνέχεια γίνεται η μέτρηση των στοιχείων, π.χ. ο συνολικός αριθμός των σελίδων όλων των δημοσιευμάτων ανά περιοδικό, ο αριθμός των βιβλιογραφικών παραπομπών ανά δημοσίευμα, ο αριθμός των δημοσιευμάτων που είναι γραμμένα από περισσότερους του ενός συγγραφείς, κτλ. Στη συνέχεια η πληροφοριομετρική έρευνα μπορεί να εφαρμόσει σ' αυτά τα δεδομένα, ένα σύνολο στατιστικών και μαθηματικών τεχνικών για διαφορετικούς κάθε φορά σκοπούς που εμπίπτουν στον ευρύτερο χώρο του πεδίου αυτού.

#### **4. Εφαρμογές της βιβλιομετρικής στο Web - Webometrics**

Υπάρχει ένα αυξανόμενο ενδιαφέρον για την εφαρμογή των τεχνικών του ευρύτερου πεδίου της βιβλιομετρικής και στο Διαδίκτυο, με την ενσωμάτωση στοιχείων και εργαλείων από πεδία όπως η επιστήμη της πληροφορικής, οι επιστήμες της επικοινωνίας, η στατιστική και η επιστήμη των υπολογιστών.

Από τα μέσα του '90 άρχισαν οι προσπάθειες εξερεύνησης της φύσης και των περιεχομένων του Web, με την εφαρμογή νέων μεθοδολογιών στα περιεχόμενα των κειμένων, τη δομή των συνδέσεων και τις μηχανές αναζήτησης. Το πεδίο αυτό άρχισε να αναπτύσσεται όταν διαπιστώθηκε πως οι ποσοτικές μέθοδοι που είχαν σχεδιαστεί για τη βιβλιομετρική ανάλυση των βιβλιογραφικών παραπομπών των άρθρων που δημοσιεύονται σε επιστημονικά περιοδικά, μπορούσαν να εφαρμοστούν και στο Web με τη χρήση εμπορικών μηχανών αναζήτησης για την παροχή δεδομένων. Αυτό ξεκίνησε αφ' ότου στη διεπαφή αναζήτησης της Alta Vista δόθηκε η δυνατότητα σύνθετων αναζητήσεων με τη χρήση τελεστών Boolean και τη δυνατότητα μελέτης των ιδιοτήτων των συνδέσεων και των URLs. (Thelwall, et al, 2003).

Αυτό το πεδίο της έρευνας ονομάζεται "Webometrics". Οι Almind & Ingwersen (1997) εργαζόμενοι για την σύγκριση μηχανών αναζήτησης του διαδικτύου με τις βάσεις δεδομένων του Institute for Scientific Information (ISI), ως πηγές βιβλιομετρικών στοιχείων, δημιούργησαν τον όρο webometrics για να ορίσουν τη ποσοτική μελέτη των φαινομένων που σχετίζονται με το Web.

Η webometrics ορίζεται ως ο τομέας που χρησιμοποιεί τις βιβλιομετρικές τεχνικές για τη μελέτη της σχέσης των διαφόρων σελίδων στον παγκόσμιο ιστό. Τέτοιες τεχνικές χρησιμοποιούν και η πληροφοριομετρική και η επιστημονομετρική. Μπορούν να χρησιμοποιηθούν για την χαρτογράφηση περιοχών του Web που φαίνεται να είναι πιο "χρήσιμες". Οι τεχνικές αυτές βασίζονται στην καταγραφή του αριθμού των συνδέσεων αυτών των σελίδων από άλλες σελίδες, τα στατιστικά χαρακτηριστικά των εγγράφων του Web και των συνδέσεων των υπερκειμένων τους, (Palmquist, 2004) και συμπεριλαμβάνουν και την ανάλυση των log files του Web και την έρευνα στις δομές των συνδέσεων και στην ανάπτυξη του Web.

#### **5. Μέθοδοι συλλογής δεδομένων στο Web**

Η συλλογή δεδομένων από το Web είναι πλέον πολύ διαδεδομένη και δυναμική. Για την συλλογή δεδομένων που αφορούν τεκμήρια του Web, δικτυακούς τόπους ή την δομή μερών του Web, χρησιμοποιούνται υπάρχοντα εργαλεία ανάκτησης πληροφοριών που είναι κυρίως οι μηχανές αναζήτησης οι οποίες βέβαια απέχουν πολύ από το να είναι τέλειες. Οι μελέτες που υπάρχουν βασίζονται σε έρευνες, συνεντεύξεις και ανάλυση συμβάντων (log analysis).

##### **5. 1. Έρευνες**

Από τα πρώτα βήματα του Web εμφανίστηκαν έρευνες για τον καθορισμό της δημογραφίας των χρηστών του, τη συλλογή πληροφοριών που αφορούν τη χρήση της τεχνολογίας του, τη χρήση των λιστών συζήτησης (discussion lists) ή την κατανόηση της χρήσης του για επιστημονικούς σκοπούς. Ουσιαστικά όλες αυτές οι έρευνες στοχεύουν στο να διαπιστώσουν ποιοι είναι οι χρήστες αλλά και τι θέλουν αυτοί ή πως κινούνται μέσα στο Web (Bar, et al, 2002).

Η έρευνα μπορεί να γίνει με τη χρήση ποσοτικών μεθόδων, όπως με ερωτηματολόγια που στέλνονται είτε με τον παραδοσιακό τρόπο είτε μέσω ηλεκτρονικού ταχυδρομείου, είτε και με συνδυασμό των δύο μεθόδων. Ένας άλλος τρόπος έρευνας είναι η δημιουργία ειδικού χώρου στο site όπου οι επισκέπτες του μπορούν να συμπληρώσουν τα στοιχεία που τους ζητούνται. Άλλες μέθοδοι που έχουν χρησιμοποιηθεί είναι οι ανοικτές συνεντεύξεις ή η ανάλυση του περιεχομένου συγκεκριμένων δικτυακών τόπων.

Συλλογή στοιχείων για έρευνες γίνεται και με τη χρήση λογισμικών και την εκτενή στατιστική ανάλυσή τους. Με τη χρήση μορφών (forms) HTML είναι δυνατή η συλλογή ποιοτικών πληροφοριών που μπορούν να χρησιμοποιηθούν σε συνδυασμό με μεγάλο όγκο ποσοτικών δεδομένων. Μεταξύ άλλων μπορούν να χρησιμοποιηθούν τεχνικές όπως η υπερδειγματοληψία και η τυχαία δειγματοληψία, για να επιτευχθεί μεγαλύτερη ακρίβεια αποτελεσμάτων. Οι έρευνες αυτές αναζητούν τις επιταγές του νόμου του Zipf, προσπαθώντας να καθορίσουν τις προτιμήσεις διαφόρων ομάδων και να συμβάλουν στην βελτίωση των διεπαφών του Web (Thelwall, et al, 2003).

## **5. 2. Έλεγχος και ανάλυση των αρχείων καταγραφής συμβάντων (log files)**

Ένας άλλος τρόπος συλλογής δεδομένων, που αφορούν και πάλι τη χρήση διαφόρων υπηρεσιών του Web, με τον έλεγχο και την ανάλυση των log files των χρηστών του Web. Στην περίπτωση αυτή μελετώνται οι ακολουθούμενες διαδικασίες, η διάρκειά τους, ο τύπος των υπηρεσιών που χρησιμοποιούνται και ο έλεγχος των ερωτήσεων που υποβλήθηκαν σε διάφορες μηχανές αναζήτησης (Bar, et al, 2002).

Με τη χρήση τεχνικών προγραμματισμού μπορούν να συλλεχθούν logs από τους web servers. Ένα πρόβλημα είναι πως οι web servers καταγράφουν την κάθε πρόσβαση ως ξεχωριστή και δεν καταγράφουν τα μονοπάτια (paths) των πηγών που ακολούθησαν οι χρήστες για να φτάσουν στη συγκεκριμένη πηγή. Αυτό στο συμβατικό περιβάλλον θα σήμαινε πως μετρούνται οι παραπομπές χωρίς να γνωρίζουμε τον αριθμό των πηγών από τις οποίες προέρχονται. Ένα άλλο σημαντικό πρόβλημα είναι πως οι ευρέσεις (hits) σε έναν server δεν αντανακλούν πραγματική χρήση, αφού πολλοί χρήστες ίσως να μην χρησιμοποιούν μια πηγή όταν η σύνδεσή της απαιτεί μεγάλο χρόνο ή διακόπτεται και οδηγούνται στη χρήση άλλων ανακτηθέντων δεδομένων. Παρόλα τα προβλήματα με ένα συνδυασμό server logs με τον κατάλληλο σχεδιασμό των πηγών του συστήματος, μπορούν να συλλεχθούν πιο αξιόπιστα δεδομένα (Thelwall, et al, 2003).

Σε έναν τυπικό Web server UNIX, υπάρχουν 4 πιθανά είδη logs:

- Αυτά που δημιουργούνται από τον ίδιο τον server
- Αυτά που δημιουργούνται μέσω του firewall (Proxy-based logs) ή μέσω κάποιου συστήματος ελέγχου πρόσβασης στον server
- Αυτά που δημιουργούνται στον client, μέσω του κώδικα εφαρμογών του client και μπορούν να καταγράφουν και να μεταφέρουν πληροφορίες στον server (π.χ. cookies)
- Αυτά που δημιουργούνται στο δίκτυο, στα προγράμματα που ελέγχουν την ασφάλεια του δικτύου και στην πρόσβαση στον server



Μερικές γενικές μέθοδοι διασφάλισης της αξιοπιστίας των αρχείων καταγραφής συμβάντων είναι οι εξής (Thelwall, et al, 2003):

- Τακτικό back up, με βάση τα χαρακτηριστικά χρήσης του server
- Αποθήκευση και της ανάλυσης των log files και των ίδιων των logs
- Δημιουργία νέων logs σε καθορισμένη χρονικά βάση
- Αποστολή των αποτελεσμάτων και των πληροφοριών που συλλέγονται από τα logs και σε άλλους ώστε να γίνεται δυνατή η σύγκρισή τους.

### **5. 3. Χρήση Crawlers**

Η συνεχής και αδιάκοπη αύξηση τεκμηρίων και ιστοσελίδων στο Web, το κάνουν τόσο μεγάλο και πολύπλοκο που η μελέτη και η απεικόνιση των χαρακτηριστικών του γίνεται όλο δυσκολότερη. Για την αναζήτηση και μελέτη μεγάλων δικτυακών τόπων μπορούν να χρησιμοποιηθούν οι γνωστές μεγάλες εμπορικές μηχανές αναζήτησης ή ένα Web crawler. Ο crawler είναι ένα λογισμικό που μπορεί αυτόματα από ένα συγκεκριμένο URL να «κατεβάσει» σελίδες και να εξάγει και να κατεβάσει από αυτές όλες τις συνδέσεις που διαθέτουν. Παρόλα αυτά δεν μπορούν να συλλέξουν το 100% των συνδέσεων γιατί για την ανεύρεση των σελίδων ακολουθούν τις συνδέσεις. Έτσι υπάρχει περίπτωση να χάσουν πολλές απομονωμένες σελίδες ή ακόμη και αυτές τις οποίες ο crawler δεν μπορεί να τις αναγνωρίσει (π.χ. JavaScript) (Thelwall, et al, 2003).

### **5. 4. Εμπορικές μηχανές αναζήτησης**

Οι μεγάλες μηχανές αναζήτησης επιλέγονται πολύ συχνά για τη συλλογή συγκεκριμένων δεδομένων από το Web. Μπορεί να συλλεχθούν δεδομένα από μία μόνο μηχανή ή να συγκριθούν δεδομένα από διάφορες μηχανές ή εργαλεία αναζήτησης. Αρκετές δίνουν τη δυνατότητα σύνθετης αναζήτησης με λογική Boolean και επιτρέπουν την αναζήτηση σε σελίδες, συνδέσεις και ολόκληρες θεματικές περιοχές με βάση όποια κριτήρια τεθούν από τον χρήστη. Παρόλα αυτά οι εμπορικές μηχανές αναζήτησης έχουν κάποια μειονεκτήματα: α) τα δεδομένα που συλλέγονται από το Web δεν μπορούν να θεωρηθούν πλήρη και ολοκληρωμένα, β) οι μηχανές αυτές δεν είναι σε θέση να ευρετηριάζουν το σύνολο των περιεχομένων του και η ακρίβεια των δεδομένων που ανακτώνται είναι πολύ συχνά αμφισβητήσιμη (Bjorneborn & Ingwersen, 2001), γ) οι αλγόριθμοι για crawling και της δημιουργίας αναφορών λειτουργίας είναι εμπορικά μυστικά με αποτέλεσμα να δημιουργούν πρόβλημα στην επιστημονική έρευνα.

### **5. 5. Δειγματοληπτική ανάκτηση**

Η έρευνα στο σύνολο του Web, λόγω του μεγάλου όγκου του, της πολυπλοκότητάς του, αλλά και λόγω τεχνικών δυσκολιών, είναι αδύνατη. Για τον λόγο αυτό όταν οι ερευνητές δεν μπορούν να αρκестούν στη χρήση ενός περιορισμένου αριθμού πηγών, καταφεύγουν στην χρήση δειγματοληπτικών μεθόδων (Bjorneborn & Ingwersen, 2001).

Κάποιες από τις έρευνες που διεξήχθησαν με τον τρόπο αυτό βασίστηκαν στη δημιουργία υποτιθέμενων «τυχαίων URLs» και/ή δείγματα λέξεων από σταθμισμένα λεξιλόγια του Web, με στόχο να μετρηθεί η κάλυψη που προσφέρουν διάφορες μηχανές αναζήτησης. Η εφαρμογή της εξαντλητικής ανάκτησης όλων των τεκμηρίων που υπάρχουν στο Web και σχετίζονται με ένα συγκεκριμένο θέμα ή μια συγκεκριμένη χώρα, επιτρέπει στους ερευνητές την δημιουργία τυχαίων δειγμάτων από το σύνολο (Bar, et al, 2002).

## **5. 6. Δειγματοληπτική ανάκτηση σελίδων**

Η ανάγκη συλλογής στοιχείων από το Web με δειγματοληπτικές μεθόδους, οδήγησε στη δημιουργία ειδικών αλγορίθμων που επιτρέπουν την ανάκτηση σελίδων, με βάση την απαιτούμενη σ' αυτές τις περιπτώσεις, τυχαία επιλογή. Η βασική προσέγγιση είναι η επιλογή σελίδων μέσα από ένα όσο το δυνατόν μεγαλύτερο αρχικό σύνολο. Ένας συνήθης τρόπος είναι η περιδιάβαση του ερευνητή σε διάφορες σελίδες εν μέρη τυχαία και εν μέρη ακολουθώντας τις συνδέσεις των σελίδων που επισκέπτεται. Σε κάθε περίπτωση πάντως ο ερευνητής πρέπει να επισκεφτεί πολύ περισσότερες σελίδες από αυτές που τελικά θα επιλέξει (Thelwall, et al, 2003).

## **5. 7. Δειγματοληπτική ανάκτηση δικτυακών τόπων**

Η μελέτη πολλών δικτυακών τόπων επιλεγμένων μέσα από ένα συγκεκριμένο και προκαθορισμένο μεγάλο σύνολο, απαιτεί την εφαρμογή κάποιας μεθοδολογίας που εξασφαλίζει το τυχαίο της επιλογής. Αν τα URLs είναι γνωστά τότε μπορεί να εφαρμοστεί οποιαδήποτε μέθοδος επιλογής τυχαίων αριθμών μέσα από ένα σύνολο, αφού για παράδειγμα πριν έχουν αριθμηθεί ή μπορεί να γίνει η επιλογή τους με βάση τις IP addresses. Υπάρχουν όμως πολλές περιπτώσεις που οι παραπάνω μέθοδοι δεν είναι δυνατόν να εφαρμοστούν. Σε κάθε περίπτωση πάντως απαιτείται η υιοθέτηση της καταλληλότερης μεθόδου που εξασφαλίζει την απαραίτητη αξιοπιστία, αλλά και απαιτεί όσο το δυνατόν λιγότερη προσφυγή σε πηγές εκτός του Web.

## **5. 8. Μελέτες μεγάλων τμημάτων του Web**

Οι μελέτες που απαιτούν την έρευνα σε πολύ μεγάλα τμήματα του Web μπορούν να χρησιμοποιήσουν τις εξής εναλλακτικές λύσεις (Turbull, 2002):

- Χρήση crawler για την ανάκτηση των ίδιων των σελίδων
- Πρόσβαση στις αναζητήσεις (crawls) όλων των σελίδων που βρίσκονται στο Internet Archive ([www.archive.org](http://www.archive.org))
- Χρήση αναζητήσεων (crawls) στο σύνολο μιας συγκεκριμένης μηχανής αναζήτησης
- Χρήση σύνθετων/εξελιγμένων αναζητήσεων σε μηχανές αναζήτησης, που ορίζουν με ακρίβεια τις σελίδες προς μέτρηση (εφαρμόζεται στην περίπτωση που δεν είναι απαραίτητη η έρευνα στο πλήρες περιεχόμενο των σελίδων)

## **6. Πληροφοριομετρική ανάλυση**

Μετά τη συλλογή δεδομένων και ανεξάρτητα από την μέθοδο συλλογής που εφαρμόστηκε, το στάδιο που ακολουθεί είναι η ανάλυσή τους. Μπορούν να χρησιμοποιηθούν απλές διαδικασίες και κάποια συγκεκριμένη στατιστική και μαθηματική ανάλυση είναι αρκετή. Σε κάποιες άλλες περιπτώσεις απαιτείται η εφαρμογή ειδικών πληροφοριομετρικών μεθόδων, μοντέλων ή νόμων που εφαρμόζονται στην ανάλυση των δεδομένων από/και για το Web.

### **6. 1. Ανάλυση βιβλιογραφικών παραπομπών (citation analysis)**

Μια από τις πρώτες μελέτες στο πεδίο αυτό αφορούσε την ανάλυση και αξιολόγηση των ηλεκτρονικών περιοδικών, μέσω της μέτρησης του αριθμού των βιβλιογραφικών παραπομπών τους. Άλλες έρευνες που ακολούθησαν ασχολήθηκαν με τις παραπομπές που είχαν οι ηλεκτρονικές πηγές (e-mails, μηνύματα σε λίστες συζήτησης, δημοσιεύσεις κάθε είδους και ηλεκτρονικές πηγές διαθέσιμες στο Διαδίκτυο). Πολύ νωρίς ακόμη οι ερευνητές παρατήρησαν πως σε μια ηλεκτρονική πηγή οι συνδέσεις που υπάρχουν θα μπορούσαν να θεωρηθούν ως το ανάλογο των βιβλιογραφικών παραπομπών στα συμβατικά δημοσιεύματα.

Μια από τις πρώτες έρευνες αυτού του είδους σε συγκεκριμένα ηλεκτρονικά περιοδικά (Smith, 1999, από Thelwall, et al, 2003), με τη χρήση της Alta Vista μέτρησε τις συνδέσεις που γινόταν από το υπόλοιπο Web. Το συμπέρασμα ήταν πως δεν υπήρχαν ομοιότητες μεταξύ των συντελεστών βαρύτητας (impact factors) όπως αυτοί ορίζονται από το Institute for Scientific Information (ISI) και όπως αυτοί συνάγονται από τη μελέτη των υπερσυνδέσεων στο Web. Δηλαδή οι υπερσυνδέσεις στα ηλεκτρονικά περιοδικά λειτουργούν διαφορετικά απ' ότι οι βιβλιογραφικές παραπομπές στο συμβατικό περιβάλλον. Μεταγενέστερες όμως έρευνες (Vaughan & Hysen, 2002, από Thelwall, et al, 2003) απέδειξαν πως υπάρχουν τελικά ομοιότητες και αυτό αποδίδεται στην ωριμότητα που απέκτησε εν τω μεταξύ το Web.

Πολλές μηχανές αναζήτησης, όπως οι Alta Vista, Hotbot και Google παρέχουν τα URLs που παραπέμπουν στις βάσεις δεδομένων τους. Αυτό δίνει την δυνατότητα εξέτασης της παροχής backlinks (σελίδες που έχουν σύνδεση προς μια συγκεκριμένη σελίδα) απ' ευθείας σε έναν δικτυακό τόπο και όχι μέσω της μηχανής αναζήτησης. Μεγάλο ενδιαφέρον παρουσιάζουν οι έρευνες που δεν μελετούν τις υπερσυνδέσεις ως ανάλογες των παραδοσιακών βιβλιογραφικών παραπομπών, αλλά επικεντρώνονται στη μελέτη τους μέσα από το σύστημα "Autonomous Citation Indexing" (ACI) που αυτόματα δημιουργεί ένα ευρετήριο παραπομπών της βιβλιογραφίας σε ηλεκτρονική μορφή. Το ACI επιτρέπει την ταυτόχρονη αναζήτηση και στα άρθρα και στις παραπομπές τους.

### **6. 2. Ανάλυση ζευγών παραπομπών και λέξεων (co-citation and co-word analysis)**

Εξετάζεται η δυνατότητα εφαρμογής των τεχνικών των ζευγών παραπομπών ή λέξεων, στο περιβάλλον του Web, στην περίπτωση που οι συνδέσεις θεωρούνται ως παραπομπές, για τον εντοπισμό και την αξιολόγηση σχετικών ή όμοιων ιστοσελίδων. Αυτό επιτυγχάνεται με τη χρήση εμπορικών μηχανών αναζήτησης και της επεξεργασίας των δεδομένων που ανακτώνται. Μελετώνται μέσω πινάκων συσχέτισης (correlation matrix) και πολυδιάστατων κλιμάκων (MDS) και τη δημιουργία ενός χάρτη αντιστοίχισης ζευγών παραπομπών (co-citation map). Άλλος τρόπος μελέτης τους είναι ειδικοί αλγόριθμοι.

### **6. 3. Ανάλυση περιεχομένου (content analysis)**

Η ανάλυση περιεχομένου των πηγών του Web και του Internet χρησιμοποιούνται για την καλύτερη κατανόηση του περιεχομένου τους. Στις διάφορες έρευνες που έχουν διεξαχθεί, έχουν αναλυθεί τα περιεχόμενα των μηνυμάτων λιστών συζήτησης, ιστοσελίδων και δικτυακών τόπων. Στις έρευνες αυτές μελετάται η χρησιμότητα του περιεχομένου των παραπάνω πεδίων, οι μέθοδοι επικοινωνίας και διάδρασης μεταξύ των συμμετεχόντων, η σχετικότητα των παραπομπών, τα χαρακτηριστικά των περιεχομένων, η δομή τους, κτλ.

Η απλή μέτρηση και η ανάλυση περιεχομένου των ιστοσελίδων μοιάζει πολύ με την ανάλυση των παραδοσιακών έντυπων δημοσιευμάτων, αφού οι συνδέσεις προς και από μια ιστοσελίδα μπορούν να θεωρηθούν και να μελετηθούν όπως οι συμβατικές παραπομπές. Όμως το Web λόγω της δυναμικής και κατανεμημένης φύσης του, συχνά διαθέτει σελίδες που παρουσιάζουν ταυτόχρονες συνδέσεις μεταξύ τους, κάτι που εκ των πραγμάτων δε μπορούσε να συμβεί στα έντυπα. Η κάλυψη από τις μηχανές αναζήτησης του συνόλου του Web, μπορεί να ερευνηθεί με τον ίδιο τρόπο όπως η κάλυψη όλων των επιστημονικών τομέων από τις βάσεις δεδομένων των βιβλιογραφικών παραπομπών. Μια άλλη διαφορά της εφαρμογής βιβλιομετρικών μεθόδων στο Web, είναι πως ενώ στα παραδοσιακά επιστημονικά δημοσιεύματα

υπάρχει μια εκ των προτέρων αξιολόγησή τους, αυτό στο Web εμφανίζεται σε πολύ μικρό βαθμό, γεγονός που αυτομάτως μειώνει την ποιότητα του περιεχομένου του.

Ένα πολύ βασικό χαρακτηριστικό του Web που το κάνει να διαφέρει από τα παραδοσιακά περιβάλλοντα δημοσίευσης είναι ο χρόνος. Ο χρόνος στο Web δίνει μια διαφορετική διάσταση στη δημοσίευση αφού σε ελάχιστο χρόνο μπορεί κανείς να δημοσιεύσει την εργασία του, αλλά και πολύ συχνά υλικό που έχει καταχωρηθεί, παύει να υφίσταται.

#### **6. 4. Αξιολόγηση με τη χρήση υπαρχόντων / νέων μεθόδων**

Μελετάται η αποτελεσματικότητα, της σχετικότητας και της ακρίβειας της ανάκτησης πληροφοριών από τις μηχανές αναζήτησης στο Web, όπως αυτά κρίνονται με τον παραδοσιακό τρόπο από τους ίδιους τους χρήστες που έκαναν την αναζήτηση. Αξιολογούνται δηλαδή οι ίδιες οι μηχανές αναζήτησης. Προστίθενται όμως και νέοι τρόποι μέτρησής τους, όπως η μέθοδος “PageRank”, που βασίζεται στην κλασική μέθοδο της ανάλυσης των παραπομπών, αλλά αντί απλώς να μετράει τον αριθμό των συνδέσεων προς ένα συγκεκριμένο τεκμήριο, υπολογίζει και την ποιότητα του τεκμηρίου το οποίο προσφέρει τη σύνδεση αυτή. Ακόμη προτείνεται η αξιολόγηση της ποιότητας των μηχανών αναζήτησης που μπορεί να βασίζεται στις συνδέσεις που γίνονται προς αυτήν ή στον αριθμό των επισκεπτών της και το πεδίο από το οποίο αυτοί προέρχονται.

#### **6. 5. Προσδιορισμός και υπολογισμός δεικτών**

Στα έντυπα περιοδικά για να υπολογιστεί η επίδραση που ασκεί ένα συγκεκριμένο περιοδικό, άρα και η φήμη ή η ποιότητά του, υπολογίζονται οι συντελεστές βαρύτητας (impact factors). Συντελεστής βαρύτητας ενός περιοδικού είναι το αριθμητικό δεδομένο που προέρχεται από τη διαίρεση του συνολικού αριθμού των παραπομπών που έχουν γίνει προς τα άρθρα που έχουν δημοσιευτεί σ’ αυτό, δια του αριθμού του συνόλου των άρθρων που έχουν δημοσιευτεί σ’ αυτό, δηλ.:

*n αριθμός παραπομπών προς το περιοδικό / n αριθμός δημοσιευμένων άρθρων στο περιοδικό)*

Ορίζονται ειδικοί δείκτες και για το Web. Ο Ingwerson το 1998 προσδιόρισε το Web Impact Factor (WIF) ως εξής (Bar, et al, 2001):

- με τις ιστοσελίδες που έχουν σύνδεσμο σ’ αυτό τον δικτυακό τόπο ή στην χώρα
- τον αριθμό των ιστοσελίδων του δικτυακού τόπου ή της χώρας

Όπως και ο κλασικός συντελεστής βαρύτητας των περιοδικών, το WIF ενός συγκεκριμένου δικτυακού τόπου ή μιας χώρας δείχνει το ποσό δυνατότητας εντοπισμού και χρήσης της στο Web. Ο συντελεστής βαρύτητας ενός δικτυακού τόπου είναι η αναλογία του αριθμού συνδέσεων που παρέχονται προς ένα δικτυακό τόπο, προς τον αριθμό των ιστοσελίδων αυτού του δικτυακού τόπου. Μάλιστα ο Rousseau (1997) χρησιμοποιεί τον όρο “sitiation” και ασχολείται με την ιστομετρική ανάλυση των παραπομπών και αναλύει τα επαναλαμβανόμενα χαρακτηριστικά κατανομής των δικτυακών τόπων και τις εισερχόμενες συνδέσεις. Όσον αφορά τη χρήση των WIFs, έχουν γίνει οι εξής παρατηρήσεις (Bjorneborn & Ingwersen, 2001):

- Όπως συμβαίνει και με τις παραπομπές των εντύπων πηγών, οι συνδέσεις των δικτυακών τόπων μπορεί να γίνονται για διαφορετικούς λόγους. Έτσι, οι συνδέσεις είναι ένα ατελές μέτρο της γενικής επιρροής ενός δικτυακού χώρου σε ολόκληρο το Web.

- Οι συμβατικοί συντελεστές βαρύτητας βασίζονται στην έννοια του «άρθρου», το οποίο είναι μια πολύ καλά καθορισμένη μονάδα. Μια ιστοσελίδα, όμως δεν καθορίζεται πολύ εύκολα και ποικίλει ανάλογα με τα πεδία πληροφοριών που περιέχουν πολλές πληροφορίες.
- Το URL δεν είναι απαραίτητο να δείχνει τη θέση μιας ιστοσελίδας.
- Οι βάσεις δεδομένων των μηχανών αναζήτησης δεν είναι απαραίτητα καλά ενημερωμένες.

## 7. Μοντέλα πληροφοριομετρικής ανάλυσης στο Web

Ο Rousseau διατύπωσε πως ο βιβλιομετρικοί νόμοι είναι εφαρμόσιμοι και στο Web. Προσάρμοσε τον νόμο του Lotka και στον αριθμό των δεδομένων των ανακτόμενων σελίδων ανά δικτυακό τόπο και στον αριθμό των παραπομπών σε ένα δικτυακό τόπο. Ο νόμος του Bradford έδειξε πως είναι εφαρμόσιμος στον προσδιορισμό βασικών ομάδων πληροφόρησης (newsgroups). Οι Hubert, [et al] έδειξαν πως η συμπεριφορά των χρηστών του Web ακολουθεί τον νόμο της διασποράς του Zipf.

### 7. 1. Τοπολογικά μοντέλα και εξόριση του Web (web mining)

Στην θεωρία των γραφικών παραστάσεων, το γράφημα είναι μια μαθηματική αναπαράσταση του δικτύου που αποτελείται από συνδεδεμένους κόμβους (nodes). Οι κόμβοι αυτοί μπορεί να είναι ανθρώπινοι (σε κοινωνικά δίκτυα), τεκμήρια (σε ένα δίκτυο παραπομπών) ή ιδέες (σε έναν θησαυρό) κλπ. Σε ένα άμεσο γράφημα τα σημεία τομής των κόμβων αναπαριστούν τις κατευθυντήριες σχέσεις μεταξύ των κόμβων. Το Web είναι ένα παράδειγμα άμεσης γραφικής παράστασης, όπου οι ιστοσελίδες είναι οι κόμβοι και οι υπερσυνδέσεις τα σημεία ακμής (edges) των κόμβων (Bjorneborn & Ingwersen, 2001).

Το Web είναι ένα τεράστιο δίκτυο του οποίου οι τοπολογικές πληροφορίες είναι άμεσα διαθέσιμες. Αν εξεταστεί ως ένα σύνολο σελίδων συνδεδεμένων μεταξύ τους και παραβλεφθούν οι πληροφορίες σχετικά με το περιεχόμενο των σελίδων, οι τοποθεσίες και τα URLs, τότε μπορεί να προκύψει ένα προσανατολισμένο γράφημα (directed graph). Το θέμα αυτό αφορά τη δημιουργία μοντέλων και του καθορισμού των επαναλαμβανόμενων χαρακτηριστικών (patterns) του.

### 7. 2. Μοντέλα σύνδεσης

Πολύ σημαντικό στοιχείο στην έρευνα για την αξιολόγηση και την κατανόηση του Web, αποτελούν τα δεδομένα και η συμπεριφορά που παρουσιάζεται στον τρόπο που δημιουργούνται οι συνδέσεις. Η έρευνα για να καταλήξει σε συγκεκριμένα ποσοτικά συμπεράσματα καταγράφει τον αριθμό των συνδέσεων προς και από ένα δικτυακό τόπο, τις ιστοσελίδες που διαθέτει, τις επισκέψεις που δέχεται ή τις σελίδες που δέχονται επισκέψεις σε ένα δικτυακό τόπο. Έχει διαπιστωθεί πως ένα χαρακτηριστικό που ισχύει είναι η «συνεχής ανάπτυξη και η προνομιακή σύνδεση», δηλαδή ο πλούσιος δικτυακός τόπος έχει την τάση να γίνεται πλουσιότερος (Barabasi & Albert, 1999, από M.Thelwall, et al, 2003).

### 7. 3. Δίκτυα “Small-Worlds”

Μια πολύ ενδιαφέρουσα διάσταση της τοπολογίας του Web προσφέρουν τα δίκτυα “small-worlds”. Στα δίκτυα επιτυγχάνεται ένα πολύ μικρό ποσοστό από συνδέσεις που ορίζονται ως «ο σύντομος δρόμος» και συνδέονται με «μακρινά» μέρη του δικτύου. Έχει αποδειχθεί ότι η τοπολογία των σχημάτων των δικτύων “small-worlds”, ισχύει στα βιολογικά, τεχνολογικά και κοινωνικά δίκτυα (Bjorneborn &

Ingwersen, 2001). Αυτό που υποστηρίζεται, είναι πως «ακόμη και ένας μικρός αριθμός συνδέσεων είναι συνήθως αρκετός για να πάμε από οποιοδήποτε σημείο οπουδήποτε, με έναν πολύ μικρό αριθμό διαδικασιών» (Berners-Lee & Cailliau, 1990, από M.Thelwall, et al, 2003).

Η τοπολογία δικτύων “small-worlds” ασχολείται με θέματα όπως η δυνατότητα πλοήγησης και πρόσβασης στην πληροφορία μέσα σε δίκτυα με πολλά τεκμήρια, με τις δυνατότητες άμεσων και σύντομων συνδέσεων, καθώς και την τεχνική συγκομιδής (harvesting) στο Web. Έρευνες διεξάγονται για τα φαινόμενα των δικτύων “small-worlds” και τις επιδράσεις τους σε μεγαλύτερα δίκτυα, όπως το WWW, σε βάσεις δεδομένων βιβλιογραφικών παραπομπών, σε εννοιολογικά δίκτυα κτλ.

#### **7. 4. Εγκάρσιες συνδέσεις (Transversal links)**

Οι εγκάρσιες συνδέσεις αναφέρονται στις σύντομες διαδρομές (shortcuts) μεταξύ ετερογενών συστοιχιών (clusters) του Web. Οι συστοιχίες του Web αποτελούνται από σχετικές διασυνδέσεις ιστοσελίδων και δικτυακών τόπων, που αντανακλούν τις θεματικές απαιτήσεις των ενδιαφερομένων κοινοτήτων. Το Web εξερευνάται με παραδοσιακό, μη αυτόματο τρόπο, ή με συγκεκριμένο λογισμικό, ακολουθώντας τις συνδέσεις από ιστοσελίδα σε ιστοσελίδα και έτσι οδηγούμαστε από μια συστοιχία σε μια άλλη πιο απομακρυσμένη, χρησιμοποιώντας μία εγκάρσια σύνδεση. Η κατανόηση της δομής των συνδέσεων έχει παίξει καθοριστικό ρόλο στην ανάπτυξη των υπερκειμενικών πληροφοριακών συστημάτων, συμπεριλαμβανομένου και του Web (Bjorneborn & Ingwersen, 2001: 74).

#### **7. 5. Μεθοδολογικές εκτιμήσεις**

Προσπαθώντας να εντοπίσουμε τις εγκάρσιες συνδέσεις στο Web, τις εξετάζουμε σε συνάρτηση με τα δεδομένα. Πιθανότατα οι περισσότερες εξωτερικές συνδέσεις επικεντρώνονται στις πιο δημοφιλείς ιστοσελίδες και σ’ αυτές που είναι εξειδικευμένες σε μια θεματική περιοχή.

Στη webometrics τα προβλήματα των μεθοδολογιών σχετίζονται με τη συλλογή πρωτογενών δεδομένων από το Web που χρησιμοποιούνται ως βάση για εμπειρικές έρευνες. Τα δευτερογενή δεδομένα που συγκεντρώνονται από τις εμπορικές μηχανές αναζήτησης, θεωρούνται μη αξιόπιστα αφού υπάρχει μεγάλη αβεβαιότητα για την κάλυψη, τη συχνότητα της ενημέρωσης των ιστοσελίδων κτλ. Εναλλακτική λύση είναι η απευθείας αναφορά σε πρωτογενή δεδομένα του Διαδικτύου. Μια πιθανή μέθοδος είναι η επονομαζόμενη «τυχαίες βόλτες» (random walks) που χρησιμοποιεί τη θεωρία των γραφικών παραστάσεων. Για να εντοπιστούν οι εγκάρσιες συνδέσεις σε μια γραφική παράσταση του Web, πρέπει να γίνουν επαρκείς «τυχαίες βόλτες» ακολουθώντας τις συνδέσεις με τυχαίο τρόπο από ιστοσελίδα σε ιστοσελίδα. Με την αναλυτική μελέτη των μονοπατιών των συνδέσεων που ακολουθήθηκαν, μπορεί να εντοπιστούν οι εγκάρσιες συνδέσεις (Bjorneborn & Ingwersen, 2001).

#### **7. 6. Ανάλυση ακολουθούμενων μονοπατιών/διαδρομών (path analysis)**

Με την μελέτη των μονοπατιών που ακολουθήθηκαν από τον ερευνητή για να μελετήσει τις συνδέσεις μεταξύ σελίδων, δικτυακών τόπων ή σε κάποια βάση δεδομένων βιβλιογραφικών παραπομπών, μπορεί να ερευνηθεί ο τρόπος με τον οποίο μπορεί κάποιος να ξεκινήσει από ένα θεματικό πεδίο και φτάνει σε ένα άλλο, π.χ. να ξεκινήσει από την μόλυνση του περιβάλλοντος και να καταλήξει στην ιατρική αντιμετώπιση των αναπνευστικών προβλημάτων. Η προσέγγιση αποκαλύπτει

στοιχεία που επιτρέπουν τη μελέτη της συμπεριφοράς των ερευνητών και τις συσχετίσεις μεταξύ ομοειδών αλλά και ετερογενών θεματικών πεδίων, στοιχεία που οδηγούν στην αποσαφήνιση των επιστημονικών πεδίων και την ανακάλυψη νέων πεδίων επιστημονικής έρευνας (Bjorneborn & Ingwersen, 2001).

## 8. Συμπεράσματα

Το Web είναι μια ανεξερεύνητη περιοχή δημοσίευσης και ανεύρεσης πληροφοριών. Η συλλογή δεδομένων για την κατανόηση των αναγκών και συμπεριφορών των χρηστών, μπορεί να οδηγήσει στην δημιουργία λειτουργικών διεπαφών χρήσης, εργαλείων, μεθόδων και πρακτικών, για τη δημοσίευση καταλληλότερου και περισσότερο αξιολογημένου περιεχομένου και εν τέλη στην ωρίμανση του μέσου. Η εφαρμογή βιβλιομετρικών και πληροφοριομετρικών μεθόδων και εργαλείων στο Web φαίνεται πως μπορεί να προσφέρει την ανάλογη βάση για έρευνα. Η webometrics είναι ένα νέο ερευνητικό πεδίο που εκμεταλλεύεται τα μέσα αυτά και προσπαθεί να αναπτύξει και νέες θεωρίες και μεθοδολογίες για την παροχή κατανόησης της τοπολογίας, των λειτουργιών και των δυνατοτήτων του.

## Βιβλιογραφία

1. Bar, Judit, Ilan Blyma and Bluma C. Peritz (2002), Informetric theories and methods for exploring the internet: an analytical survey of recent research literature, *Library Trends*, 50, 3 (371-392).
2. Bjorneborn, Lennart and Peter Ingwersen (2001), Perspective of webometrics, *Scientometrics*, 50, 1 (65-82).
3. Boudourides, Moses A., Beatrice Sigrist and Philippos D. Alevizos. *Webometrics and the Self-Organization of the European Information Society*, <http://hyperion.math.upatras.gr/webometrics/>, [πρόσβαση 24-7-2007].
4. Casserly, Mary F. and James E. Bird (2003) Web citation availability: analysis and implications for scholarship, *College & Research Libraries*, 64, 4 (300-317).
5. Larson, Ray R. (1996) *Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of the cyberspace*, University of California, <http://sherlock.berkeley.edu/asis96/asis96.html>, [πρόσβαση 24-7-2007].
6. Palmquist, Ruth A. *Bibliometrics*, University of Texas, <http://www.gslis.utexas.edu/~palmquis/courses/biblio.html> [πρόσβαση 24-7-2007].
7. Rousseau, Ronald (2002) Journal evaluation: technical and practical issues *Library Trends*, 50, 3 (418-439).
8. Thelwall, Mike, Liwen Vaughan and Lennart Bjorneborn (2005) Webometrics, *Annual Review of Information Science and Technology (ARIST)*, 39 (81-135).
9. Turnbull, Don. *Bibliometrics and the World-Wide Web*, University of Toronto, <http://www.ischool.utexas.edu/~donturn/research/bibweb.html> [πρόσβαση 24-7-2007].
10. Wilson, Concepcion N. (1999) Informetrics, *Annual Review of Information Science and Technology (ARIST)*, 34 (107-247).
11. Wormell, Irene (1998) Informetrics: an emerging subdiscipline in information science, *Asian Libraries*, 7, 10 (257-268).