

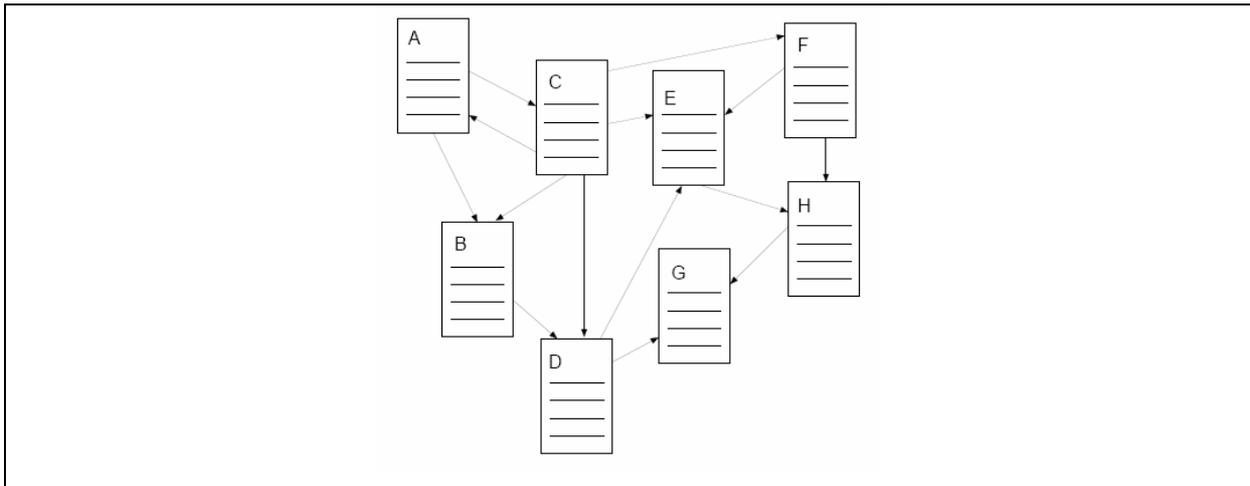
Mein Leben als Web-Robot (EG 7)

(siehe auch <http://www-student.cs.uni-bonn.de/~wichmann/writings/webcrawlers/index.html> Kapitel 3.2)

Aufgabe 5:

Versetzen Sie sich in den nächsten Minuten in die Rolle eines Web-Robots.

Angenommen Sie starten Ihre Suche im Dokument A (siehe unten), welche Seite besuchen Sie als nächste?



Gar nicht so einfach, oder? Im Prinzip haben Sie zwei systematische Vorgehensmöglichkeiten. Die erste ist die folgende:

- 1) Füge den URL (im obigen Beispiel A) der Startseite in eine Liste ein.
- 2) Ist die Liste leer? Wenn ja, dann STOP.
- 3) Nehme **den ersten URL aus der Liste**, lösche diesen aus der Liste und hole die zugehörige Web-Seite aus dem Internet.
- 4) Durchsuche die neue Web-Seite nach neuen URLs. Füge noch nicht besuchte URLs **an das Ende** der Liste an.
- 5) Gehe zu Schritt 2).

Aufgaben

- 1) In welcher Reihenfolge werden die obigen Dokumente durchlaufen?
- 2) Ändert sich die Reihenfolge, wenn man in Punkt 4 die neuen URLs nichts ans Ende, sondern an die Spitze der Liste einfügt? Wenn ja, wie?
- 3) Welche der beiden Strategien würden Sie anwenden, wenn Sie möglichst schnell **in die Tiefe** einer WWW-Site (z.B. unsere Homepage) eindringen wollen, welche wenn Sie sich einen eher **breiteren** Grobübersicht über die Inhalte der WWW-Site beschaffen wollen.