電話帳情報のハイパーテキスト化のデザイン

高橋克巳 (NTT 情報通信研究所) [1], 高田敏弘 (NTT 基礎研究所),
三浦信幸, 島健一 (NTT ソフトウエア研究所)

takahasi@isl.ntt.jp, takada@mdma.brl.ntt.jp, miura@slab.ntt.jp, kshima@slab.ntt.jp

あらまし

本論文では電話帳情報を WWW アプリケーションとして実験公開するために行なったハイパーテキスト化のデザインについて述べる．はじめにオンライン電話帳の問題点の分析を行ない，従来の並列入力インタフェイスに代わって，ユーザの希望に従って行なわれる個別入力インタフェイスが必要なこと，検索結果や過程においても関連の情報へのハイパーリンクを用意して，ユーザの情報発見を支援する工夫が必要なことを述べる．つぎにデザインへの要求条件を，情報検索とハイパーテキストの側面から考察する．前者からは《名前》《住所》《職業》を使って電話帳情報を組織化することについて述べる．後者からは検索条件の組合せで記述される電話帳情報の断片間の関係を定義し，それぞれの有用性について考察する．残りの段階では，対象となる電話帳データの分析を行ない，付加的に利用する地図，広告情報までを融合した上で，前記分析に基づいたハイパーテキスト化のデザインについて述べる．

このデザインは，ユーザの要求に従った《名前》《住所》《職業》による電話帳情報の組織化を支援し，検索結果ならびに過程に設けられた関連の情報へのハイパーリンクによって情報の発見を支援する特徴を持つ．このデザインに基づいて開発された WWW 電話帳は，データをオブジェクト指向データベースに格納し，ユーザと OODB の仲介を行ない，検索結果から動的に HTML ファイルおよびハイパーリンクを作成する CGI プログラムを持つ．本サーバは <http://www.pearnet.org/jtd/> にて公開されている．

キーワード

● 電話帳 ● ハイパーテキスト ● WWW ● 情報検索 ● オブジェクト指向データベース ● マンマシンインタフェイス

Designing a hypertext telephone directory

Katsumi Takahashi (NTT Information and Communication System Laboratories) [2],
Toshihiro Takada (NTT NTT Basic Research Laboratories),
Nobuyuki Miura, Ken-ichi Shima (NTT Software Laboratories),

takahasi@isl.ntt.jp, takada@mdma.brl.ntt.jp, miura@slab.ntt.jp, kshima@slab.ntt.jp

abstract

This paper introduces the design of a hypertext telephone directory for the WWW application. This user interface provides a organizing method for the directory information by “names”, “addresses”, and “categories”. It also supports users for information finding by hyperlinks dynamically created on results or processes of the retrieval. The WWW telephone directory implemented on this design consists of an object-oriented database and CGI programs. The OODB contains the phone listings. CGI program mediates between OODB and the user and processes HTML file dynamically from the search results. This server can be accessed on <http://www.pearnet.org/jtd/>.

key words

● Telephone Directory ● Hypertext ● WWW ● Information Retrieval ● OODB ● Man Machine Interface

[1] 〒 238-03 神奈川県横須賀市武 1-2356, TEL (0468)59-2024

[2] 1-2356 Take Yokosuka-shi Kanagawa 238-03 Japan, TEL +81 468 59 2024

1 はじめに

電話帳は実世界情報へのインデックスという性格を持っており，インターネットでも有効に活用する必要がある．筆者らは電話帳情報をインターネットのWWWアプリケーションとして公開，提供する実験を行なっている．本論文はこの実験で行なった電話帳情報のハイパーテキスト化のデザインの目指すものについて述べ，実装を報告する．なお，デザインの方法論はハイパーメディアのデザイン法RMM [4]，OOHDM[5] などを参考にしている．

本実験 Japan Telephone Directory (JTD) は，WWWを使ったマルチメディア電話帳システムである．URLは <http://www.pearnet.org/jtd/>，1995年10月より実験を開始した [6].

2 問題点の分析

2.1 電話帳のオンライン化

図1にこれまでのオンライン電話帳の実験について示す．

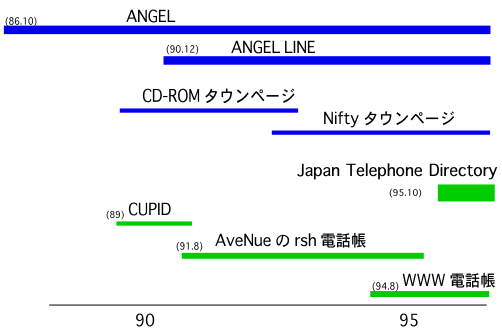

図 1: オンライン電話帳のあゆみ

ANGELは104番号案内に用いられるシステムで，ここから電話帳情報のオンライン検索が始まった．ANGEL LINEはANGELのデータベースをパソコン通信で一般利用者のアクセスを可能としたものである．JTD以下の4項目は研究目的で行なわれた実験である．CUPIDはパソコンとホストを電話回線でつないだ，利用者直接利用型実験で，最初の付加価値電話帳の実験である．その結果，総合的な検索時間の短縮が課題になった [1]．AveNueのrsh電話帳 [2] は，ユーザ制限をせず，コマンドの制限のみをする特殊なrsh[3] を使った電話帳情報提供の実験である．この実験では研究者の日常に密着した情報提供に成功た．アクセスが容易になったため番号を調べる以外の利用法も報告された[3].

これらの先行実験からわかった問題点は，

- 親しみやすいインタフェイス
  - 効果的な対話法 (質問文生成の支援／情報発見の支援)
  - 問い合わせのゆれや誤りの克服
- 提供するメディア
  - 情報へ到達するまでのコストが最小限であること
- 付加価値をつけること
  - 地図，広告などのメディアを有効に利用すること

などである．

2.2 入力インタフェイスについて

代表的な検索の入力インタフェイスにはメニュー選択，項目入力，自然語入力である [13]．メニュー選択はキー入力を減らし，意思決定の構造をはっきりさせる利点があるが，自由度が低かったり，項目が多くなり過ぎる欠点がある．項目入力と自然語入力は表現の自由度が高いが，入力にあいまいさが生ずる場合がある．

パソコン通信を使った番号案内ANGEL-LINEや，フランスのミニテルの番号案内などで一般的に使われているインタフェイスは項目入力で，ユーザは必要に応じて名前，住所など複数の項目に並列的に入力を行なうことができる．このインタフェイスには次の2つの問題点がある．(1) 入力項目が多ければ多いほど，入力誤りが起こる可能性が高くなり，またユーザがそれを発見することも難しくなる．(2) この形式がユーザに不要な入力をしいる可能性がある．この並列的に入力できる形式では，ユーザはできるだけ多くの情報を入力してしまう傾向があることが報告されている [7]．しかし，例えば日常会話などで「銀行を特定するとき，銀行はどこにでもあるから名前や住所が必要であるが，動物園はそれだけでおおむね特定できてしまう」，という性質があるように，全ての条件は必ずしも全ての検索において必須ではく，必須でない条件の入力は，ゆれやあいまいさを引き起こし，ユーザの手間やストレスとなる．

104で知られる番号案内の検索インタフェイスも上記の並列入力式であるが，番号案内のオペレータはエキスパートとしての経験からここで述べた問題点をうまく回避し，最高のパフォーマンスを上げている．しかしWWWユーザのようなシステムに知識のない一般ユーザにとってはこの形式は得策でないという結論に達する．むしろ，ユーザが望む検索条件だけをユーザの望むタイミングで，かつ必要最小限にとどめて受けつける個別入力インタフェイスが必要と考えられる．

[3] たとえば難しい固有名詞のかな漢字変換として使用された．

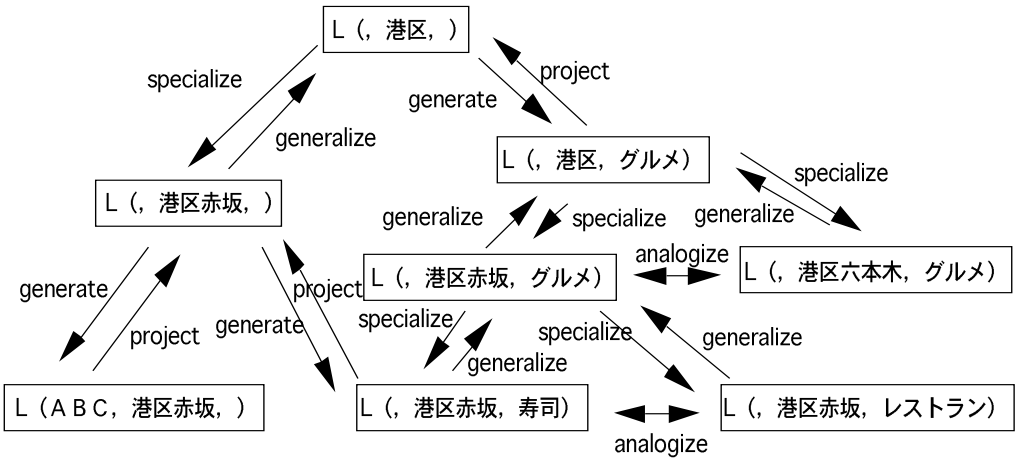

図 2: 掲載集合の関係図

紙の電話帳，あるいは現在のオンラインの電話帳システムの多くは，

始まり→目的の明確化
→検索 (ページをめくる) →番号を得る
→終り

という繰り返しから成り立っている．しかし多くの WWW サーバは，このように特定の目的を達成するためだけでなく，数々のハイパーリンクをユーザにたどらせて，ユーザに情報を発見させることを狙ったものが多い．WWW アプリケーションとして電話帳をとらえると，(1) 検索結果からの情報の発見の支援 (検索結果に関連した情報へのリンク)，(2) 複数の検索方法の提供 (名前／住所と職業／地図からなど) などが必要と考えられる．また (3) 従来たいくつなものと考えられていた，検索過程にも工夫を凝らし，情報発見の手がかりとなるような設計が必要と考えられる．また (4) 電話帳は文字情報が中心であるが，写真入りの広告などいわゆるマルチメディア情報がより求められると考えられる．

3 情報の組織化とナビゲーション

本章では，電話帳システムを，情報検索とハイパーテキストのそれぞれの立場から考察し，デザインのための要求条件の整理を行なう．

3.1 情報の組織化－情報検索の立場から－

電話帳システムを情報検索の立場から分析する．電話帳情報は伝統的に，《名前》《住所》《職業》《電話番号》という項目を持っており，補助的に，《広告》《地図》を持っているものもある．

電話帳検索は，ユーザが主に《名前》《住所》《職業》を使って電話帳情報を組織化することの手助けをすることが中心的課題である．

入力は自然語，キーワード，メニュー選択などが考えられるが，自由度とあいまいさの扱いのトレードオフが問題である．住所は地図の入力が適している．

なおこの 3 つの要素に (創立年などの) 《時間》と (従業員数などの) 《連続量》を加えると，スマート・イエローページ (Pacific Bell の紙のイエローページ) のデザインで知られるワーマンの情報の 5 つの情報の組織化方法 [8] になる．

3.2 情報のナビゲーション－ハイパーテキストの立場から－

電話帳情報をハイパーテキストの立場から分析する．電話帳情報には広くは画像広告や地図なども含まれるが，それらのハイパーテキスト化はインプリメンテーションの部分で触れることとし，ここでは《名前》《住所》《職業》《電話番号》からなる「電話帳掲載」もしくはその集まりを単位としたハイパーテキスト化について考察する．

ある掲載に「関連する」掲載は無数に考えられるが，掲載を記述する項目に注目すると次の関係がわかりやすい．(1) 名前が同じ／似ているもの，(2) 住所が同じ／似ているもの，(3) 職業が同じ／似ているもの．同じようにハイパーテキスト化の単位を「掲載の集まり」に広げて考てみる．ユーザが掲載を組織化できる方法を前節で考えた《名前》《住所》《職業》であると仮定する (それぞれを $N(ame)$，$A(dress)$，$C(ategory)$ とする)．N, A, C で決まる状態 $L(N,A,C)$ があり，これは掲載の集合である．(例えば L(”ABC 軒”，”東京都新宿区”，”ラーメン屋”) など．L(””，”東京都渋谷区”，””)(= 渋谷区の掲載全て) などとしてもよい．) これらの集合の項目 N, A, C に対し

て，次の関係を定義する (図2).

(1)[analogy] 意味あるいは表記が近いものとの置き換え．(例)”赤坂 ”と ”六本木 ”，”銀行 ”と ”保険 ”，”shochiku”と ”syoutiku”．(2)[generation] 条件を指定すること．(3)[projection] 条件を削除すること．(4)[specialization] 条件の縮小．(例)”港区 ”→ ”港区赤坂 ”，”グルメ ”→ ”フレンチレストラン ”．(5)[generalization] 条件の拡大．Address, Category に対して．(例)specialization の逆.

4 電話帳データの分析

この章では，実験データを分析し名前，住所，職業の情報量と検索インタフェイスの関係について調べる．実験に用いるデータは CitySource1995 年版に掲載されている東京 23 区，大阪市，京都市の電話帳である．データベースに登録されているレコード総数は約 3 万，表記は全て英語 (ローマ字) である．

検索は次のようにモデル化する．データベースに登録されているレコード数を N．検索条件の入力に対して，1 回の検索の出力レコード数には上限の正数 n があり，それが n レコード以下に絞り込まれたとき検索レコードの出力が行なわれるとする．検索条件は，name, category, address の 3 条件の入力とする．

このときある検索条件で表される掲載の集合をを L とし，L に該当するレコードがデータベース中に m 個存在する $(m > 0)$ とする．L に該当するレコードがみつかる確率 p は m/N だから，L の情報量 $I(L)$ は次の式で表すことができる．この量は L が m レコード以下に絞り込むための下限値である.

$$I(L) = -\log_2 \frac{m}{N}$$

また，エントロピー H は次の式で与えられる.

$$H = -\sum_i p_i \log p_i$$

検索のさいに出力が許されるレコード数の上限を n=100 とすると，検索が成立するための下限の情報量は N=30000 であることから 8.2(bit) である.

ここで name を検索条件としたときを考えよう．図3 はデータベースの name フィールドに存在する単語 (1 単語を検索条件とした時) の情報量を単語の出現頻度順に表したものである．異なり単語数は 16,000 である.

情報量が検索下限情報量 (8.2 bit) に達しない単語は ltd, International, Bank など 100 種類ほどしかない．すなわちしばしば企業名に使われる 100 種の単語以外を検索条件として指定した場合には，一単語入力だけで検索が成立することがわかる．次に category または address を検索条件としたときを考えよう．本データベースでは category として職業分類は約 600 通り，address として区名は約 70 通りあるが，職業分類 1 つを検索条件とした時のエントロピーは 7.9(bit)，同じく区名は 4.9(bit) である．このことは，特に区名では検索が成立しない場合もあることを示している．しかしこれら 2 つを「千代田区の書店」のように組み合わせて検索条件を作ると，情報量が検索下限情報量に達しない組合せは 12 通りしか存在しない (表 1).

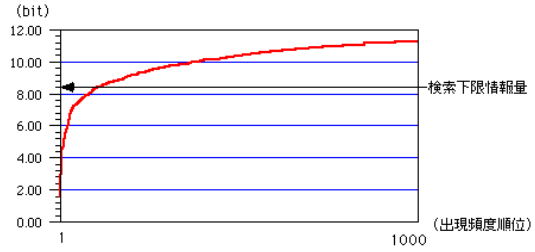

図 3: 企業名の単語の出現頻度順位と情報量

情報量	職業名	区名
5.86	TRADING COMPANIES	Osaka Chuo-ku
7.01	RESTAURANTS	Tokyo Minato-ku
7.38	TRADING COMPANIES	Osaka Kita-ku
7.43	TRADING COMPANIES	Tokyo Chiyoda-ku
7.54	TRADING COMPANIES	Tokyo Minato-ku
7.60	TRADING COMPANIES	Tokyo Chuo-ku
7.61	TRADING COMPANIES	Osaka Nishi-ku
7.65	ASSOCIATIONS	Tokyo Chiyoda-ku
7.69	BANKS	Tokyo Chiyoda-ku
7.85	ASSOCIATIONS	Tokyo Minato-ku
7.87	TRAVEL AGENCIES	Tokyo Minato-ku
8.01	RESTAURANTS OSAKA	Osaka Chuo-ku
8.23	PATENT ATTORNEYS	Tokyo Minato-ku
8.23	BANKS	Tokyo Chuo-ku
8.24	CHEMICALS	Tokyo Chiyoda-ku

表 1: 情報量の少ない職業名と区名の組合せ

この結果から，「名前 1 単語」，または「区名と職業」を入力するだけでほとんどの場合検索が成立すると予測できる.

5 ハイパーテキスト化のデザイン

以上の議論を元にして行なったデザインの概要を示す.

- 名前，住所，職業の 3 つを検索条件とする.

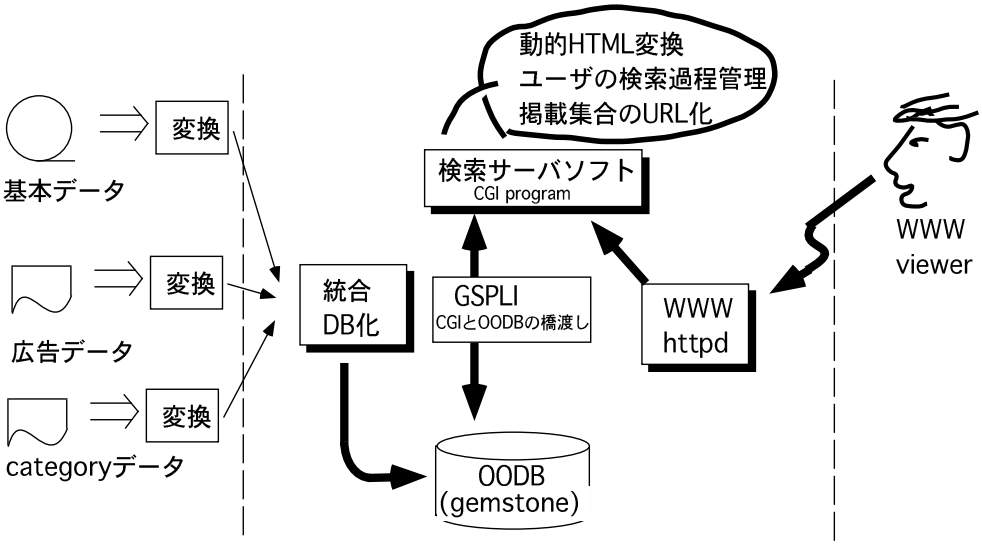

図 4: システム構成

- 各検索条件の同時入力を禁ずる.
- 任意の時点で任意の検索条件の入力／変更を可能にする.
- 可能な限り選択メニューを用意する.
- 文字列入力をさせる場合はあいまいさ対策をする.
- 検索条件で表される掲載集合をナビゲーションの単位とする.
- 掲載集合間に generation, projection, specialization, generalization, analogy の関係を与える.
- 任意の掲載集合に対して「ふさわしい」掲載集合をリンクする.

6 インプリメンテーション

図 4 に本システムの構成を示す.

- 電話帳データはオブジェクト指向データベース GemStone[9] に格納し，掲載情報および画像広告の管理を行なった [10].
- ユーザの入力をハイパーテキスト化のデザインに従って制御し，OODB と通信し，HTML を自動生成して，検索集合の URL 化を行なう検索サーバソフト (CGI) を作成した.
- 詳細地図 (建物の位置がわかる程度) および情報ページ (観光案内のドキュメント) からデータベースへのハイパーリンクを設け，他メディアとの融合を試みた.
- 入力に対し該当の掲載数が出力上限値 100 を越えない場合は掲載の出力を行ない，越える場合はユーザによる検索条件の付加 (generation) を可能とした (例えば，指定した職業が上限値を越えた場合は，住所のリストが件数つきで表示される).
- 出力結果には自動的に住所，職業からのハイパーリンクを生成し， projection を可能とした.
- 名前で "shochiku"， "syoutiku" などのようにローマ字表記のゆれのために，類似した文字列を持つ掲載を同時にかつ適合度の高い順に出力する処理を行なった (generation と analogy の混合).
- 職業のハイパーリンクには類似職業へのリンクを含み， analogy を可能とした.

7 他の動向

インターネットでは各国の電話会社を中心に電話帳情報の提供が行なわれている.

AT&T 800 Directory[4] は， WWW で提供された電話帳の草分け的存在である．単純な職業，名前の検索のほか，検索結果からの職業へのリンクなどを可能にしている． NYNEX Interactive Yellow

[4] <http://www.tollfree.att.net/>

Pages[5]は全米1,600万件のイエローページを提供するシステムであり，電話帳掲載からそれぞれの会社へのWebページへのリンクを付けることを試みている．フランスのThe French electronic phonebook[6]は良く知られた「ミニテル」へのゲートウエイを提供している．

WWWではもはや電話帳情報と，いわゆる「ホームページ」の情報との境界は明確に存在しないといえる．YAHOO[7]などのインターネットディレクトリやサイバーモールなどと呼ばれる仮想的なマーケットがそれにあたる．後者に関してはJPドメイン内でもNRI Cyber Business Park[8], Electronic Commerce Network[9], Cyber Publishing Japan[10]などが実験 / サービスを行なっている．

利用者の側にたった電話帳も存在する．個人のインターネットユーザのホームページにそれぞれの観点から作成されたレストランガイドなどは，そのわかりやすい例である．一つ一つのガイドに含まれる店の数は多くはないが，それぞれの内容は利用者の視点が貫かれており，またネットワーク内にちらばった総体は無視できないヴォリュームであると考えられる．また一定の電話帳リストを提供し，そこにユーザが自由にコメントを書き込めるサービスも存在する．例えば米国西海岸のBay Area Restaurant Guide [11]や荻窪味の店 [12]などはいち早くその機能を実現した例である．

8 おわりに

ユーザの要求に従った《名前》《住所》《職業》による電話帳情報の組織化を支援し，検索結果ならびに過程に設けられた関連の情報へのハイパーリンクによって情報の発見を支援する特徴を持つ電話帳情報のハイパーテキスト化のデザインについて述べた．

今後は，提案の検索インタフェイスの評価実験を行ない，さらに以下の項目についての検討を行なう予定である．(1) 選択肢の少ないメニューを自動生成するなどの，検索のしやすさの追求，(2)「関連のある」情報をリンクすることによる，情報発見のしやすさの追求，(3) 地図を始めとする電話帳情報の規模に見合った外部情報とのリンク方式の検討．さらには，電話帳情報は，広く多目的に利用される必要がある．インターネットでの利用をベースとして，移動体環境での情報検索 [11]，分散情報検索のための共有知識としての利用 [12] などの研究を並行して行なう予定である．

[5] <http://www.niyp.com/>

[6] <http://www.epita.fr:5000/11/>

[7] <http://www.yahoo.com/>

[8] <http://www.cyber-bp.or.jp/>

[9] <http://www.commerce.or.jp/>

[10] <http://www.toppan.co.jp/>

[11] <http://www.w3.com/bayfood/>

[12] <http://www.dec-j.co.jp/exp/south/ajinomise_home.html>

参考文献

[1] 橋田，武藤，戸部，高野．高付加価値番号案内システム (CUPID) の概要．NTT R&D Vol 39 No 6 (1990).

[2] 高橋，岩瀬．ネットワーク上の電話帳システム．44回情処全大 (1992).

[3] 梅村恭司，山崎憲一．サーバクライアントモデルを使った広域データベースの試み．41回情処全大 (1990).

[4] T. Isakowitz, E. A. Stohr, and P. Balasubramanian. RMM: A Methodology for Structured Hypermedia Design. Com.ACM Vol 38, No 8 (1995).

[5] D. Schwabe and G. Rossi. The Object-Oriented Hypermedia Design Model. Com.ACM Vol 38, No 8.

[6] 島健一，高橋克巳，三浦信幸．インターネット版マルチメディア電話帳の構築．Japan World Wide Web Conference '95. (1995)

[7] 永吉剛，岩瀬成人．統計量を用いた電話帳データベース検索方式．情処 DBS99-9. (1994)

[8] Wurman, Richiard S. INFORMATION ANXIETY. BANTAM BOOKS. pp.59 (1989)

[9] 安達隆，大嶋高行．GemStone V4.1 の概要．信学会知能ソフトウエア工学研究会．(1995-11)

[10] 島健一，高橋克巳，三浦信幸．オブジェクト指向データベースを用いたインターネット版マルチメディア電話帳．信学会データ工学研究会．(1995-12)

[11] 篠原章夫，高橋克巳，森原一郎，服部文夫，PHS & 携帯情報端末を利用した通信サービス，(情処プログラミングシンポジウム) 95.7

[12] K. Takahashi, Y. Nishibe, I. Morihara and F. Hattori. Collecting Shop and Service Information with Software Agents. Practical Applications of Intelligent Agents and Multi-Agents:PAAM'96 (1996).

[13] B. Shneiderman. Designing the Use Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley. pp. 57 (1987).