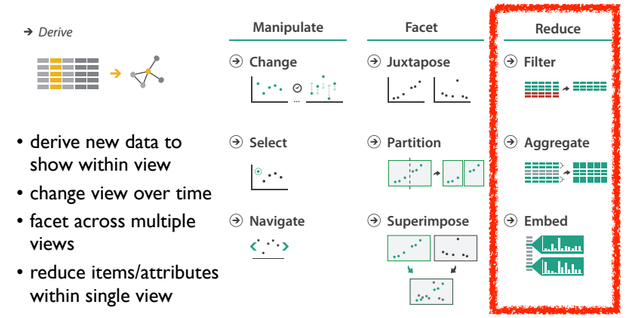
Lectures 5&6: Reduce Items & Attributes

Tamara Munzner
Department of Computer Science
University of British Columbia

DSCI 532: Data Visualization II
Lectures 5&6: 3 & 5 April 2017

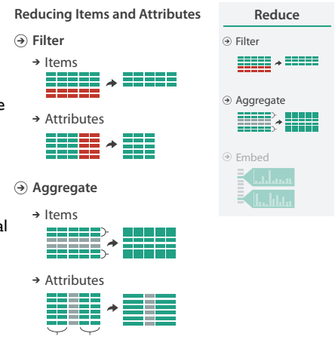
https://github.ubc.ca/ubc-mds-2016/DSCI_532_viz-2_students

How to handle complexity: 1 previous strategy + 3 more



- derive new data to show within view
- change view over time
- facet across multiple views
- reduce items/attributes within single view

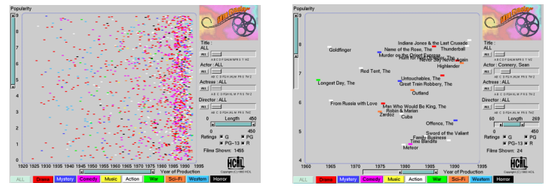
Reduce items and attributes



- reduce/increase: inverses
- filter
  - –pro: straightforward and intuitive
    - to understand and compute
  - –con: out of sight, out of mind
- aggregation
  - –pro: inform about whole set
  - –con: difficult to avoid losing signal
- not mutually exclusive
  - –combine filter, aggregate
  - –combine reduce, change, facet

Idiom: **dynamic filtering** System: **FilmFinder**

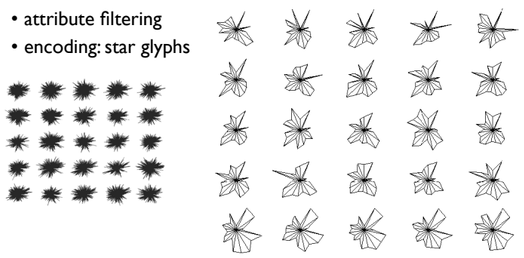
- item filtering
- browse through tightly coupled interaction
  - –alternative to queries that might return far too many or too few



[Visual information seeking:Tight coupling of dynamic query filters with starfield displays. Ahlberg and Shneiderman. Proc.ACM Conf. on Human Factors in Computing Systems (CHI), pp. 313–317, 1994.]

Idiom: **DOSFA**

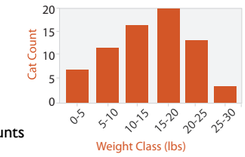
- attribute filtering
- encoding: star glyphs



[Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration Of High Dimensional Datasets. Yang, Peng,Ward, and. Rundensteiner. Proc. IEEE Symp. Information Visualization (InfoVis), pp. 105–112, 2003.]

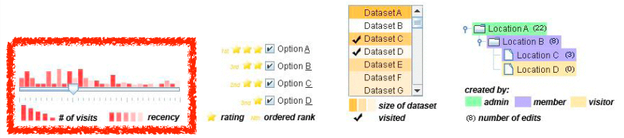
Idiom: **histogram**

- static item aggregation
- task: find distribution
- data: table
- derived data
  - –new table: keys are bins, values are counts
- bin size crucial
  - –pattern can change dramatically depending on discretization
  - –opportunity for interaction: control bin size on the fly



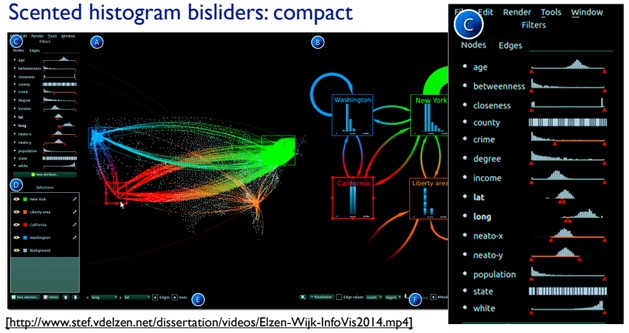
Idiom: **scented widgets**

- augment widgets for filtering to show ***information scent***
  - –cues to show whether value in drilling down further vs looking elsewhere
- concise, in part of screen normally considered control panel



[Scented Widgets: Improving Navigation Cues with Embedded Visualizations. Willett, Heer, and Agrawala. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2007) 13:6 (2007), 1129–1136.]

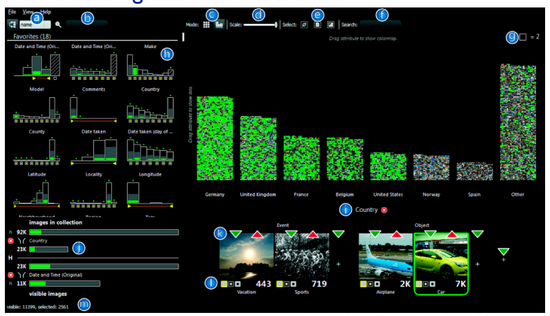
Scented histogram bisliders: compact



[http://www.stef.vdelzen.net/dissertation/videos/Elzen-Wijk-InfoVis2014.mp4]

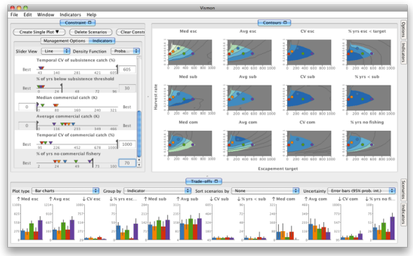
[Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. van den Elzen and van Wijk,TVCG 20(12) 2014.]

Scented histogram bisliders: detailed



[ICLIC: Interactive categorization of large image collections. van der Corput and van Wijk. Proc. PacificVis 2016.]

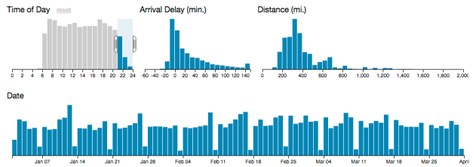
Scented histogram bisliders: details staged System: **VisMon**



http://vismon.cs.univie.ac.at/documentation.html#video

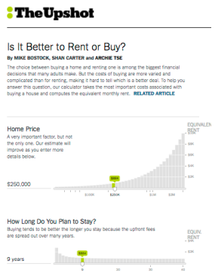
Idiom: **cross filtering** System: **Crossfilter**

- item filtering
- coordinated views/controls combined
  - all scented histogram bisliders update when any ranges change



[http://square.github.io/crossfilter/]

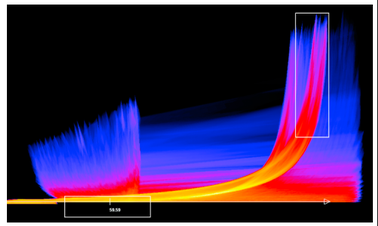
Idiom: **cross filtering**



[https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0]

Idiom: **continuous scatterplot**

- static item aggregation
- data: table
- derived data: table
  - – key attribs x,y for pixels
  - – quant attrib: overplot density
- dense space-filling 2D matrix
- color: sequential categorical hue + ordered luminance colormap

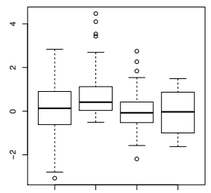


[http://www.vis.uni-stuttgart.de/~bachthsn/scatterplot/]

[Continuous Scatterplots. Bachthaler and Weiskopf. IEEE TVCG (Proc.Vis 08) 14:6 (2008), 1428–1435. 2008.]

Idiom: **boxplot**

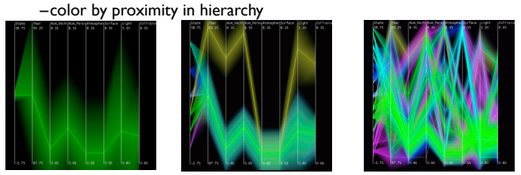
- static item aggregation
- task: find distribution
- data: table
- derived data
  - –5 quant attribs
    - median: central line
    - lower and upper quartile: boxes
    - lower upper fences: whiskers
      - –values beyond which items are outliers
  - –outliers beyond fence cutoffs explicitly shown



[40 years of boxplots. Wickham and Stryjewski. 2012. had.co.nz]

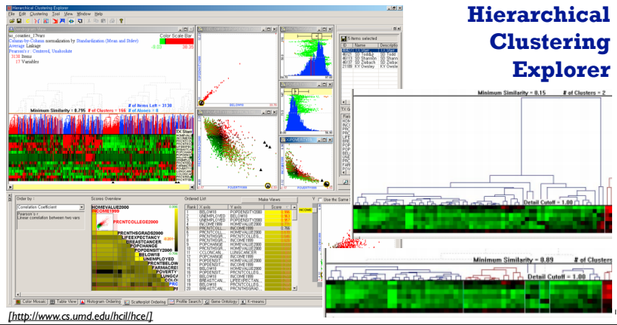
Idiom: aggregation via hierarchical clustering System: **Hierarchical parallel coordinates**

- dynamic item aggregation
- derived data: ***hierarchical clustering*** (control, not visible)
- encoding:
  - –cluster band with variable transparency, line at mean, width by min/max values
  - –color by proximity in hierarchy



[Hierarchical Parallel Coordinates for Exploration of Large Datasets. Fua,Ward, and Rundensteiner. Proc. IEEE Visualization Conference (Vis '99), pp. 43– 50, 1999.]

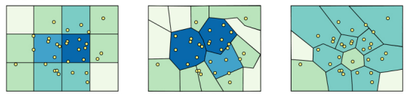
Idiom: aggregation via hierarchical clustering (visible) System: **Hierarchical Clustering Explorer**



[http://www.cs.umd.edu/hcil/hce/]

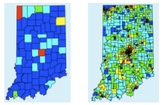
Spatial aggregation

- MAUP: Modifiable Areal Unit Problem
  - –gerrymandering (manipulating voting district boundaries) is only one example!
  - –zone effects



[http://www.e-education.psu.edu/geog486/l4_p7.html, Fig 4.cg.6]

  - –scale effects

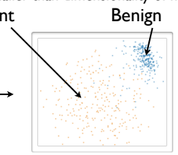


https://blog.cartographica.com/blog/2011/5/19/the-modifiable-areal-unit-problem-in-gis.html

Dimensionality reduction

- attribute aggregation
  - –derive low-dimensional target space from high-dimensional measured space
    - capture most of variance with minimal error
  - –use when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables

Tumor Measurement Data → DR →



data: 9D measured space

derived data: 2D target space

Dimensionality vs attribute reduction

- vocab use in field not consistent
  - –dimension/attribute
- attribute reduction: reduce set with filtering
  - –includes orthographic projection
- dimensionality reduction: create smaller set of new dims/attribs
  - –typically implies dimensional aggregation, not just filtering
  - –vocab: projection/mapping

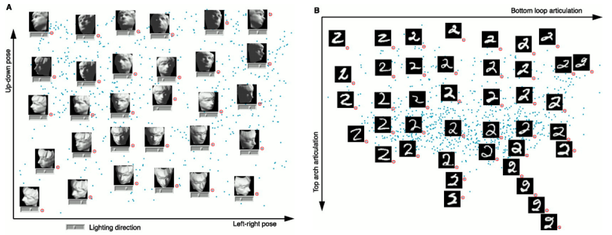
Dimensionality reduction & visualization

- why do people do DR?
  - –improve performance of downstream algorithm
    - avoid curse of dimensionality
  - –data analysis
    - if look at the output: visual data analysis
- abstract tasks when visualizing DR data
  - – dimension-oriented tasks
    - naming synthesized dims, mapping synthesized dims to original dims
  - – cluster-oriented tasks
    - verifying clusters, naming clusters, matching clusters and classes

[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Dimension-oriented tasks

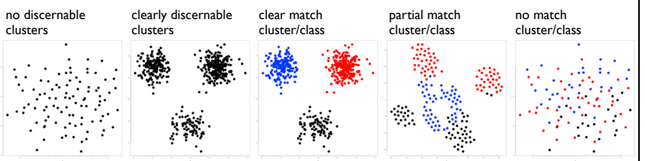
- naming synthesized dims: inspect data represented by lowD points



[A global geometric framework for nonlinear dimensionality reduction. Tenenbaum, de Silva, and Langford. Science, 290(5500):2319–2323, 2000.]

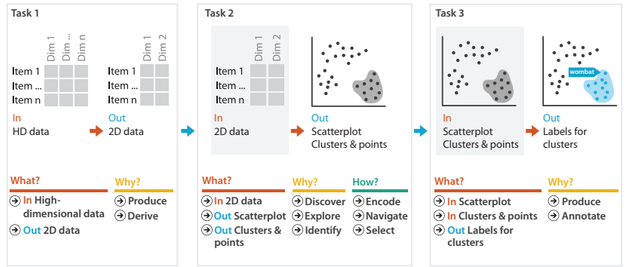
Cluster-oriented tasks

- verifying, naming, matching to classes

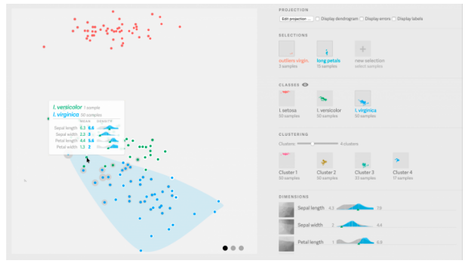


[Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. Brehmer, Sedlmair, Ingram, and Munzner. Proc. BELIV 2014.]

Idiom: **Dimensionality reduction for documents**



Interacting with dimensionally reduced data

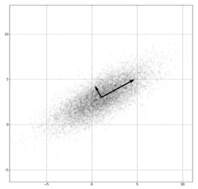


[https://uclab.fh-potsdam.de/projects/probing-projections/]

[Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions. Stahnke, Dörk, Müller, and Thom. IEEE TVCG (Proc. InfoVis 2015) 22(1):629-38 2016.]

Linear dimensionality reduction

- principal components analysis (PCA)
  - –finding axes: first with most variance, second with next most, …
  - –describe location of each point as linear combination of weights for each axis
    - mapping synthesized dims to original dims



[http://en.wikipedia.org/wiki/File:GaussianScatterPCA.png]

Nonlinear dimensionality reduction

- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attribs
  - –new dimensions often cannot be easily related to originals
    - –mapping synthesized dims to original dims task is difficult
- many techniques proposed
  - –many literatures: visualization, machine learning, optimization, psychology, ...
  - –techniques: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE,…
    - –t-SNE: excellent for clusters
      - – but some trickiness remains: http://distill.pub/2016/misread-tsne/
    - –MDS: confusingly, entire family of techniques, both linear and nonlinear
      - – minimize stress or strain metrics
      - – early formulations equivalent to PCA

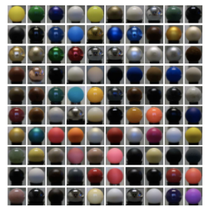
VDA with DR example: nonlinear vs linear

- DR for computer graphics reflectance model
  - –goal: simulate how light bounces off materials to make realistic pictures
    - computer graphics: BRDF (reflectance)
  - –idea: measure what light does with real materials

[Fig 2. Matusik, Pfister, Brand, and McMillan. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Capturing & using material reflectance

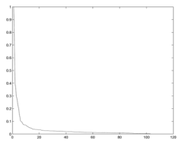
- reflectance measurement: interaction of light with real materials (spheres)
- result: 104 high-res images of material
  - –each image 4M pixels
- goal: image synthesis
  - –simulate completely new materials
- need for more concise model
  - – 104 materials * 4M pixels = 400M dims
  - –want concise model with meaningful knobs
    - how shiny/greasy/metallic
    - DR to the rescue!

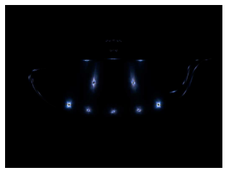


[Figs 5/6. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Linear DR

- first try: PCA (linear)
- result: error falls off sharply after ~45 dimensions
  - –scree plots: error vs number of dimensions in lowD projection
- problem: physically impossible intermediate points when simulating new materials
  - –specular highlights cannot have holes!

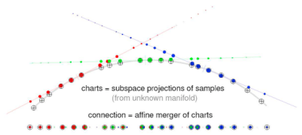


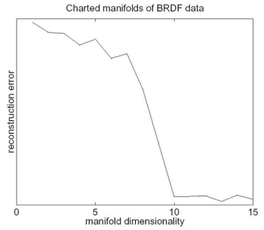


[Figs 6/7. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Nonlinear DR

- second try: charting (nonlinear DR technique)
  - –scree plot suggests 10-15 dims
  - –note: dim estimate depends on technique used!

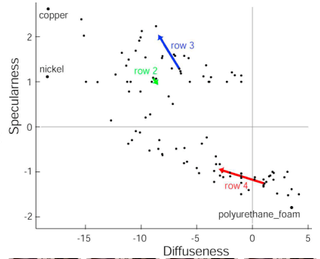




[Fig 10/11. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Finding semantics for synthetic dimensions

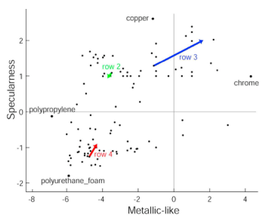
- look for meaning in scatterplots
  - –synthetic dims created by algorithm but named by human analysts
  - –points represent real-world images (spheres)
  - –people inspect images corresponding to points to decide if axis could have meaningful name
- cross-check meaning
  - –arrows show simulated images (teapots) made from model
  - –check if those match dimension semantics



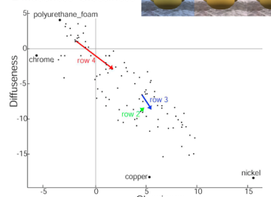
[Fig 12/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Understanding synthetic dimensions

Specular-Metallic



Diffuseness-Glossiness



[Fig 13/14/16. Matusik et al. A Data-Driven Reflectance Model. SIGGRAPH 2003]

Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *–Chap 13: Reduce Items and Attributes*
- *Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines.* Elmqvist and Fekete. IEEE Transactions on Visualization and Computer Graphics 16:3 (2010), 439–454.

Embed: Focus+Context

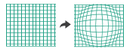
- combine information within single view
- elide
  - –selectively filter and aggregate
- superimpose layer
  - –local lens
- distortion design choices
  - –region shape: radial, rectilinear, complex
  - –how many regions: one, many
  - –region extent: local, global
  - –interaction metaphor

Embed

→ Elide Data

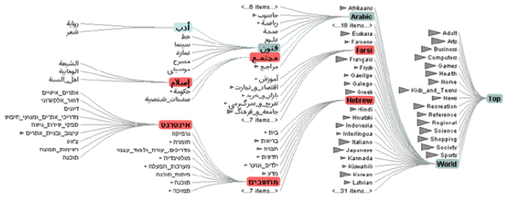
→ Superimpose Layer

→ Distort Geometry



Idiom: **DOITrees Revisited**

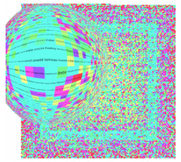
- elide
  - –some items dynamically filtered out
  - –some items dynamically aggregated together
  - –some items shown in detail

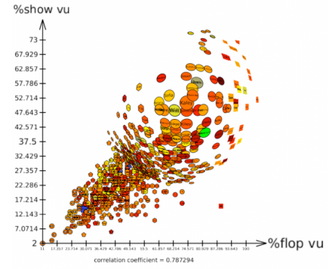


[DOITrees Revisited: Scalable, Space-Constrained Visualization of Hierarchical Data. Heer and Card. Proc. Advanced Visual Interfaces (AVI), pp. 421–424, 2004.]

Idiom: **Fisheye Lens**

- distort geometry
  - –shape: radial
  - –focus: single extent
  - –extent: local
  - –metaphor: draggable lens

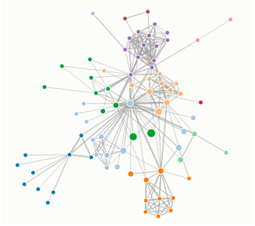




http://tulip.labri.fr/TulipDrupal/?q=node/351
http://tulip.labri.fr/TulipDrupal/?q=node/371

Idiom: **Fisheye Lens**

System: **D3**

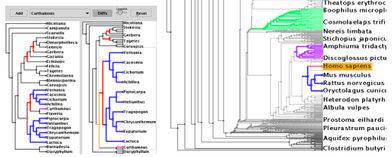


[D3 Fisheye Lens](https://bost.ocks.org/mike/fisheye/)

Idiom: **Stretch and Squish Navigation**

System: **TreeJuxtaposer**

- distort geometry
  - –shape: rectilinear
  - –foci: multiple
  - –impact: global
  - –metaphor: stretch and squish, borders fixed



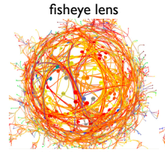
[https://youtu.be/GdaPj8a9QEo]

[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. Munzner, Guimbretiere, Tasiran, Zhang, and Zhou. ACM Transactions on Graphics (Proc. SIGGRAPH) 22:3 (2003), 453– 462.]

Distortion costs and benefits

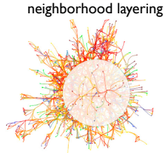
- benefits
  - –combine focus and context information in single view
- costs
  - –length comparisons impaired
    - network/tree topology comparisons unaffected: connection, containment
  - –effects of distortion unclear if original structure unfamiliar
  - –object constancy/tracking maybe impaired

fisheye lens



magnifying lens

neighborhood layering



Bring and Go

[https://www.youtube.com/watch?v=hm2oFBqVM9o]

[Living Flows: Enhanced Exploration of Edge-Bundled Graphs Based on GPU-Intensive Edge Rendering. Lambert, Auber, and Melançon. Proc. Intl. Conf. Information Visualisation (IV), pp. 523–530, 2010.]

Further reading

- Visualization Analysis and Design. Munzner. AK Peters / CRC Press, Oct 2014.
  - *–Chap 14: Embed: Focus+Context*
- *A Review of Overview+Detail, Zooming, and Focus+Context Interfaces.* Cockburn, Karlson, and Bederson. ACM Computing Surveys 41:1 (2008), 1–31.
- *A Guide to Visual Multi-Level Interface Design From Synthesis of Empirical Study Evidence.* Lam and Munzner. Synthesis Lectures on Visualization Series, Morgan Claypool, 2010.
- *Hierarchical Aggregation for Information Visualization: Overview, Techniques and Design Guidelines.* Elmqvist and Fekete. IEEE Transactions on Visualization and Computer Graphics 16:3 (2010), 439–454.
- *A Fisheye Follow-up: Further Reflection on Focus + Context.* Furnas. Proc. ACM Conf. Human Factors in Computing Systems (CHI), pp. 999–1008, 2006.