Project Report:
Semantic Portal Business and Economics

Kai Eckert
Computer Science Institute
University of Mannheim
Germany

Magnus Pfeffer
University Library
University of Mannheim
Germany

NKOS Workshop
September 19th 2008
Aarhus, Denmark

Project Goal

- Creating a OPAC+ Library Search Enginge
  - Content
    - Library media
    - All licenced fulltext documents
    - Focus on economics
  - Modern user interface
    - Thesaurus-based search and retrieval
    - Drill-down using facets
    - Support multiple thesauri

Research Topics

- Automatic indexing in the field of economics
- Thesaurus-based user search interfaces
- Multi-thesaurus indexing and search

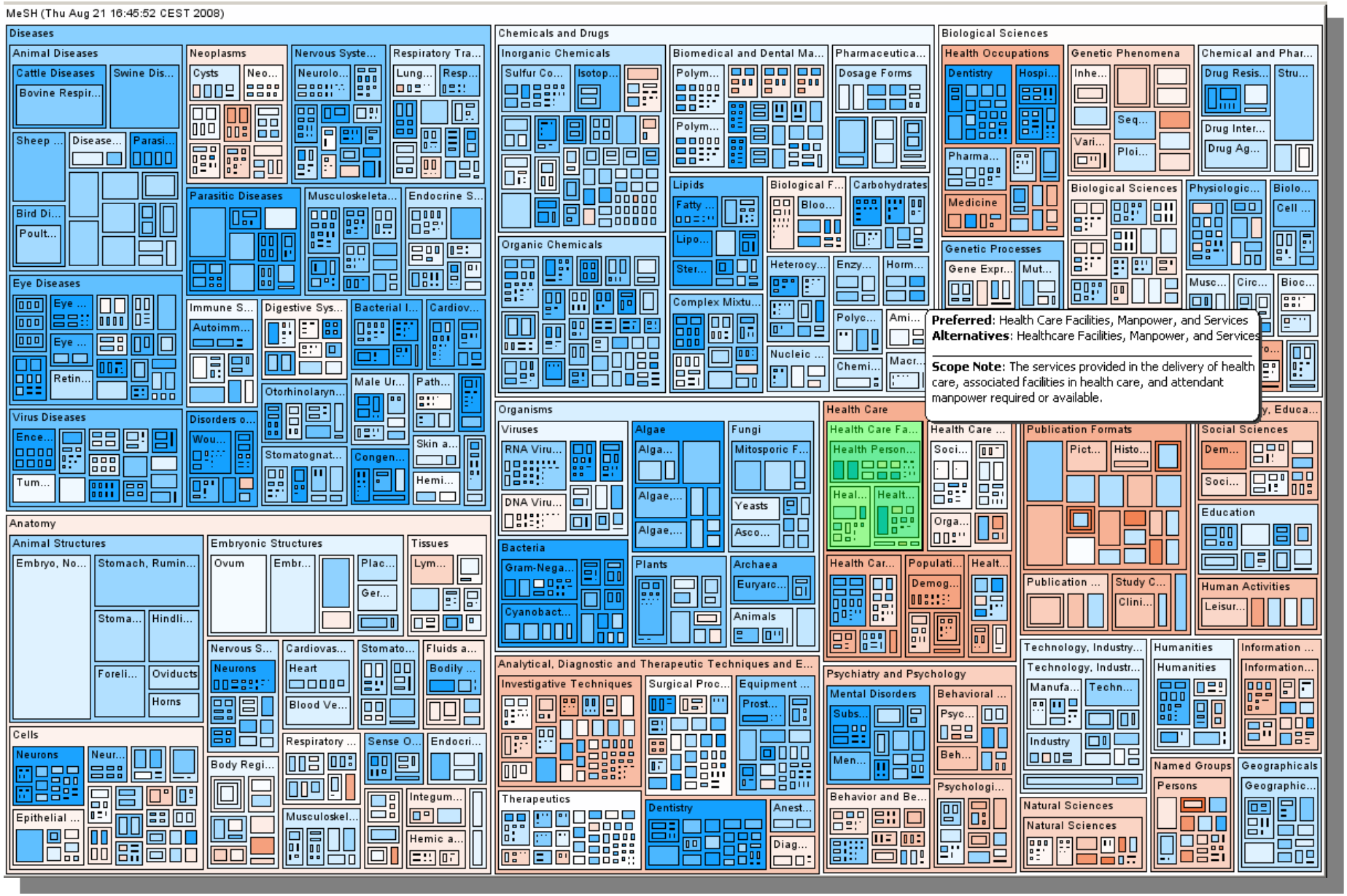
Current Status

- Prototype indexing system
  - Elsevier journal articles
  - STW Thesaurus
  - Collexis Search Engine
- Datasets
  - Automatic indexing results
  - Manually indexed articles as gold standard

Automatic Indexing Assessment

- Precision and recall comparison
  - Meaningless numbers on the macro level
  - Tedious on the micro level
- Visual analysis using Semtinel
  - Per concept IC-Diff analysis
  - Treemap for navigation
  - Easy identification of critical concepts

IC Diff Analysis with Semtinel



Automatic Indexing Assessment cont.

- Editing of example critical thesaurus concepts
  - Lack of sysnonyms
  - Insufficient disamgibuation
  - Overly broad concepts
- Reindexing
  - Improved Precision and recall

Further Steps

- Analysis and Semtinel Tool
  - Improve framework (SKOS loader)
  - Document based analysis methods
- Multi-Thesaurus Retrieval
  - Multiple indexes
  - Merging multiple thesauri
  - UI Design

Further Steps cont.

- Prototype retrieval system
  - Collexis engine and user interface
  - User study
- Integration into library systems
  - Representation using RDF and DC
  - Evaluation of Ex Libris “Primo” product

Open Questions

- How can one judge indexing results? Is our approach reasonable?
- More ideas or use-cases for Semtinel? Feature-Requests? (e.g. Ontology-Editor, ...)

Thank you for your attention.

kai@informatik.uni-mannheim.de

magnus.pfeffer@bib.uni-mannheim.de

Additional Slides

IC Diff Analysis

Information Content:

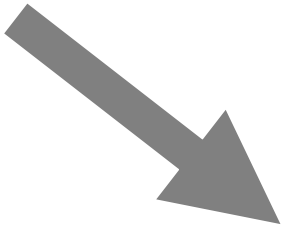
- Proposed by Resnik
- Depends on Frequency in Document Base

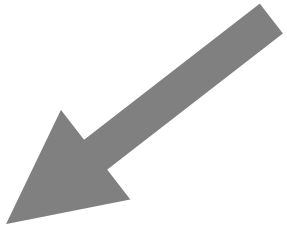
$$IC(c) = -\log P(c)$$

Intrinsic Information Content:

- Proposed by Seco, Veale und Hayes
- Based on the Number of Subconcepts

$$IIC(c) = -\log\left(\frac{hypo(c)+1}{max}\right)$$

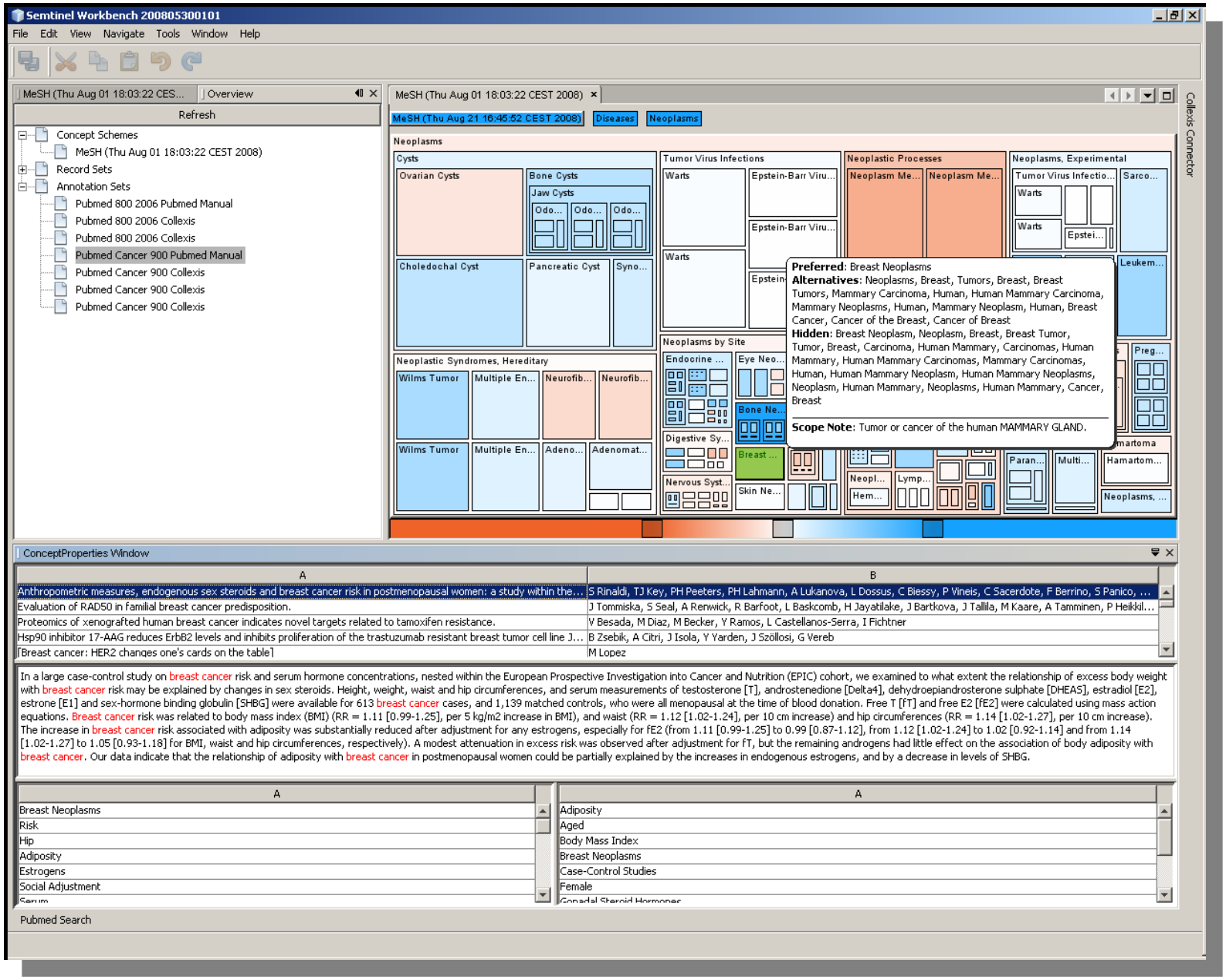




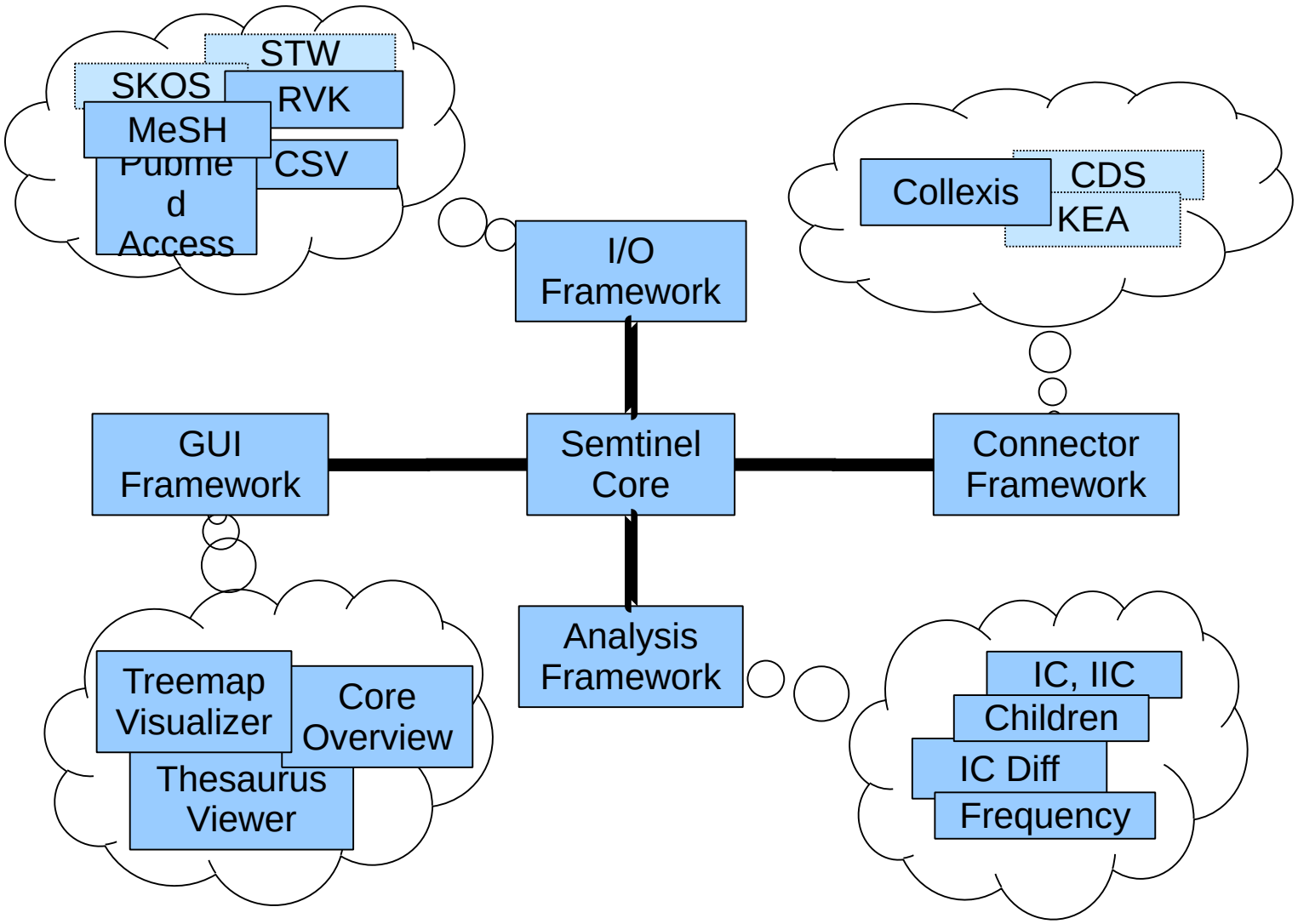
$$D_{IC}(c) = IC(c) - IIC(c)$$

Intuitive: A value between *-1* and *1* that says, if a concept has a suspicious frequency regarding its position in the thesaurus.

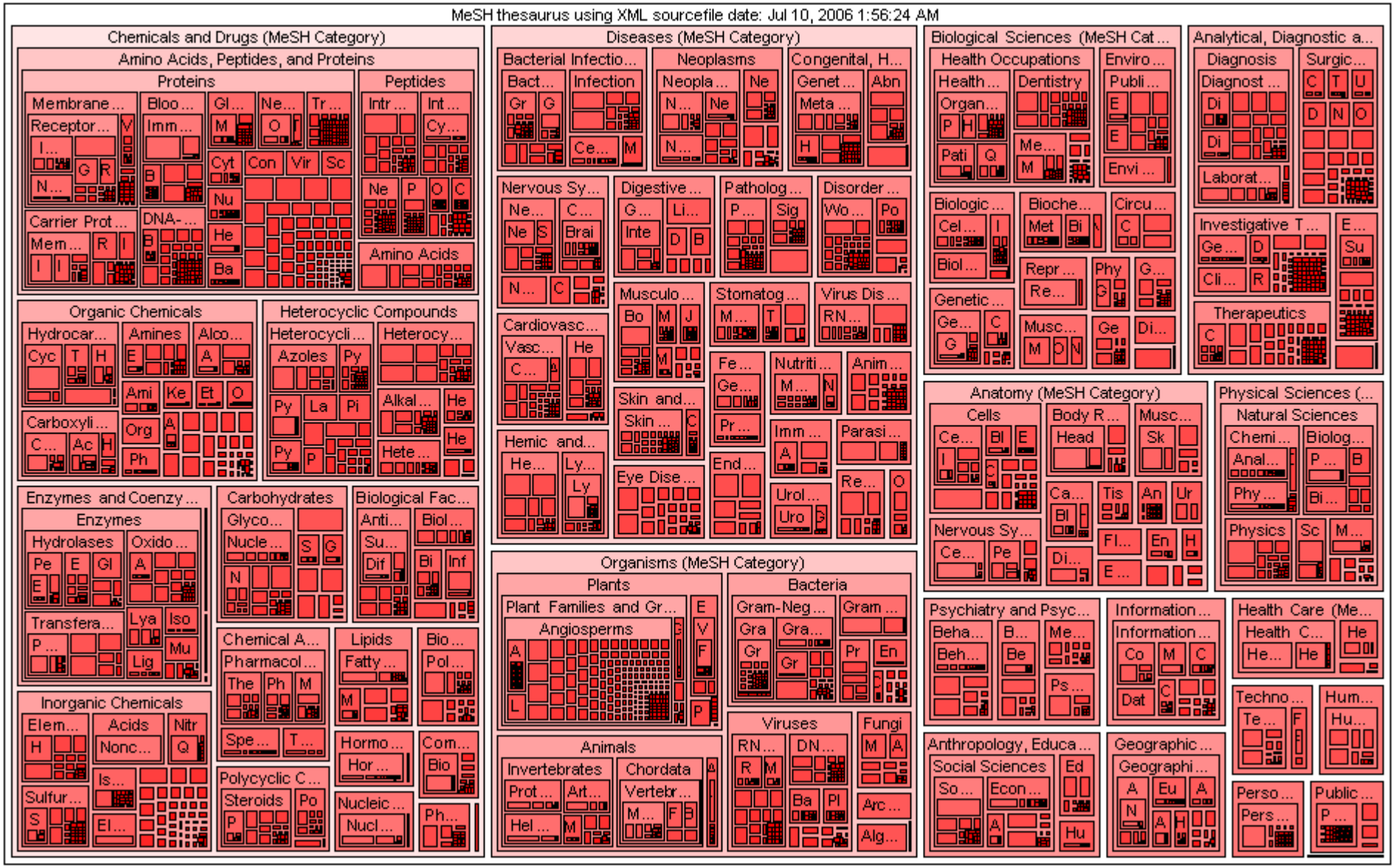
Semtinel Workbench



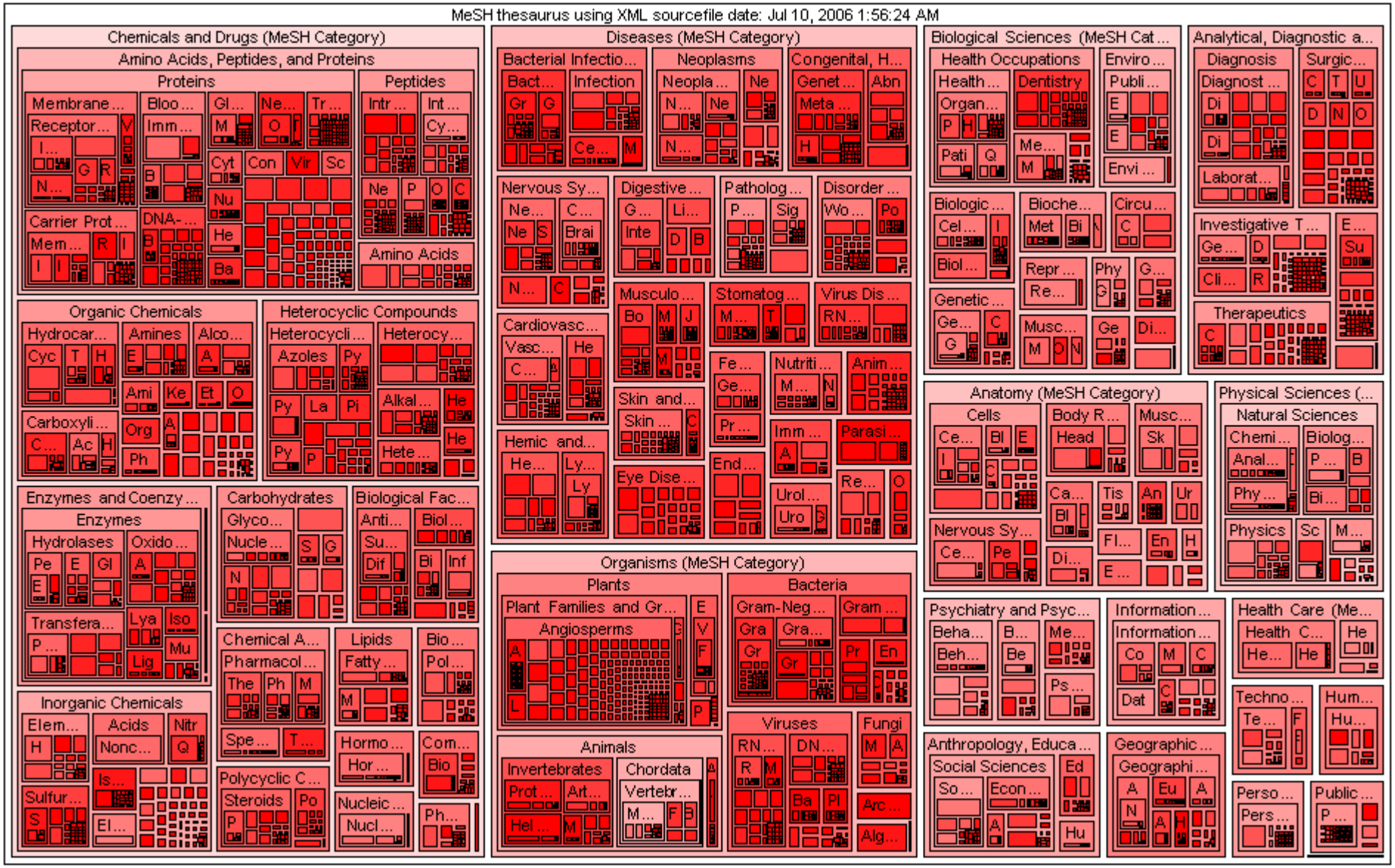
Semtinel API



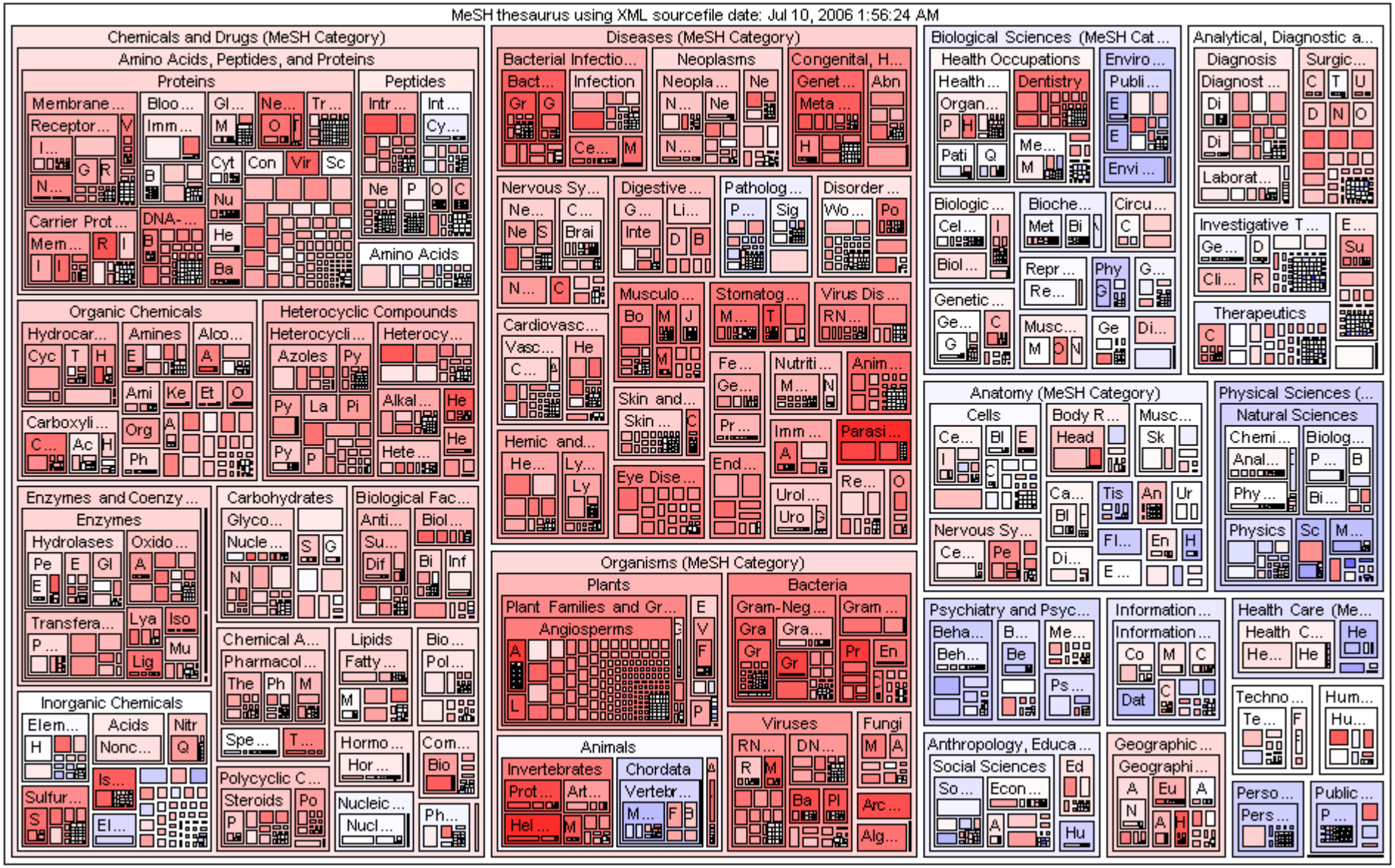
Intrinsic Information Content



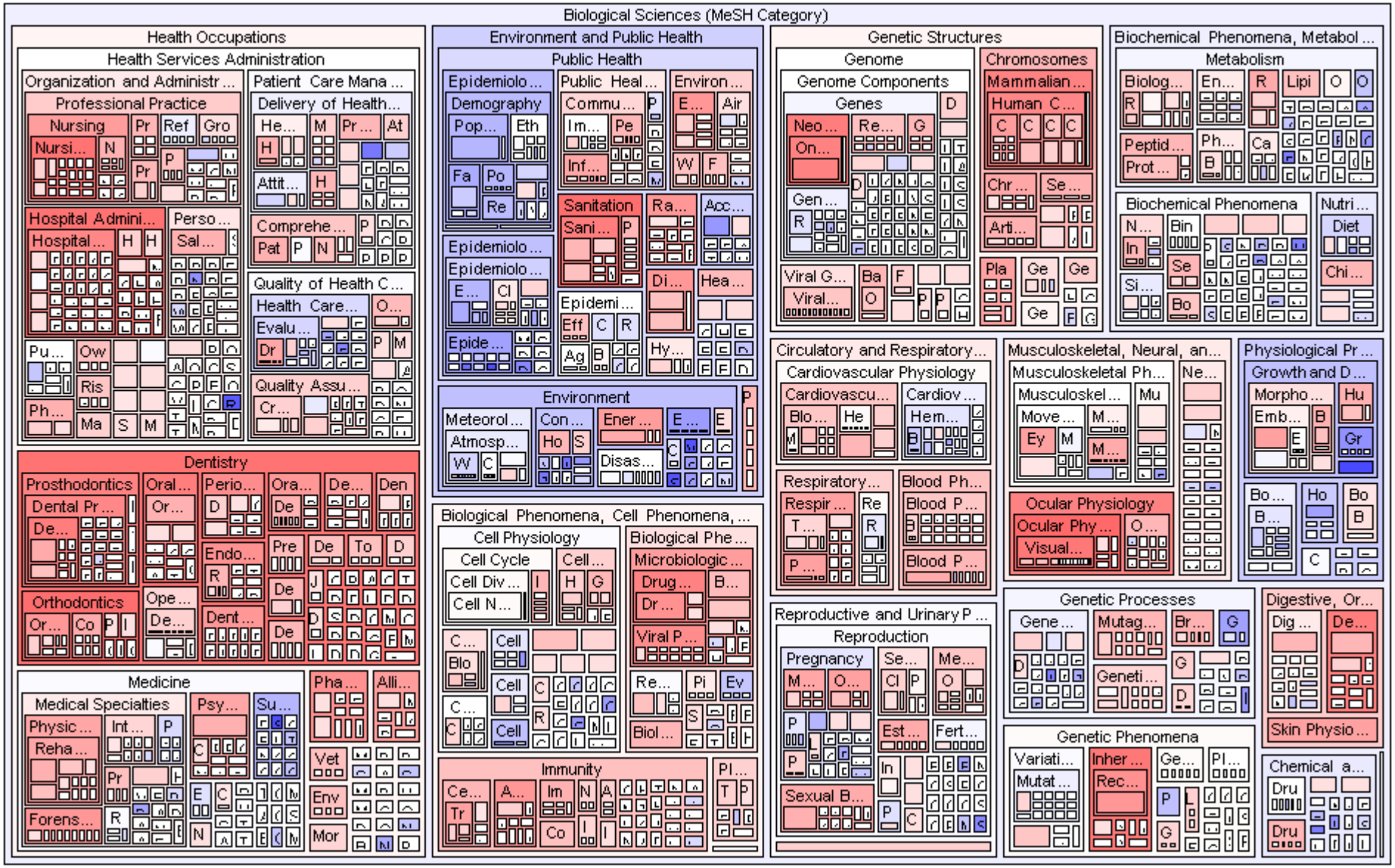
Information Content



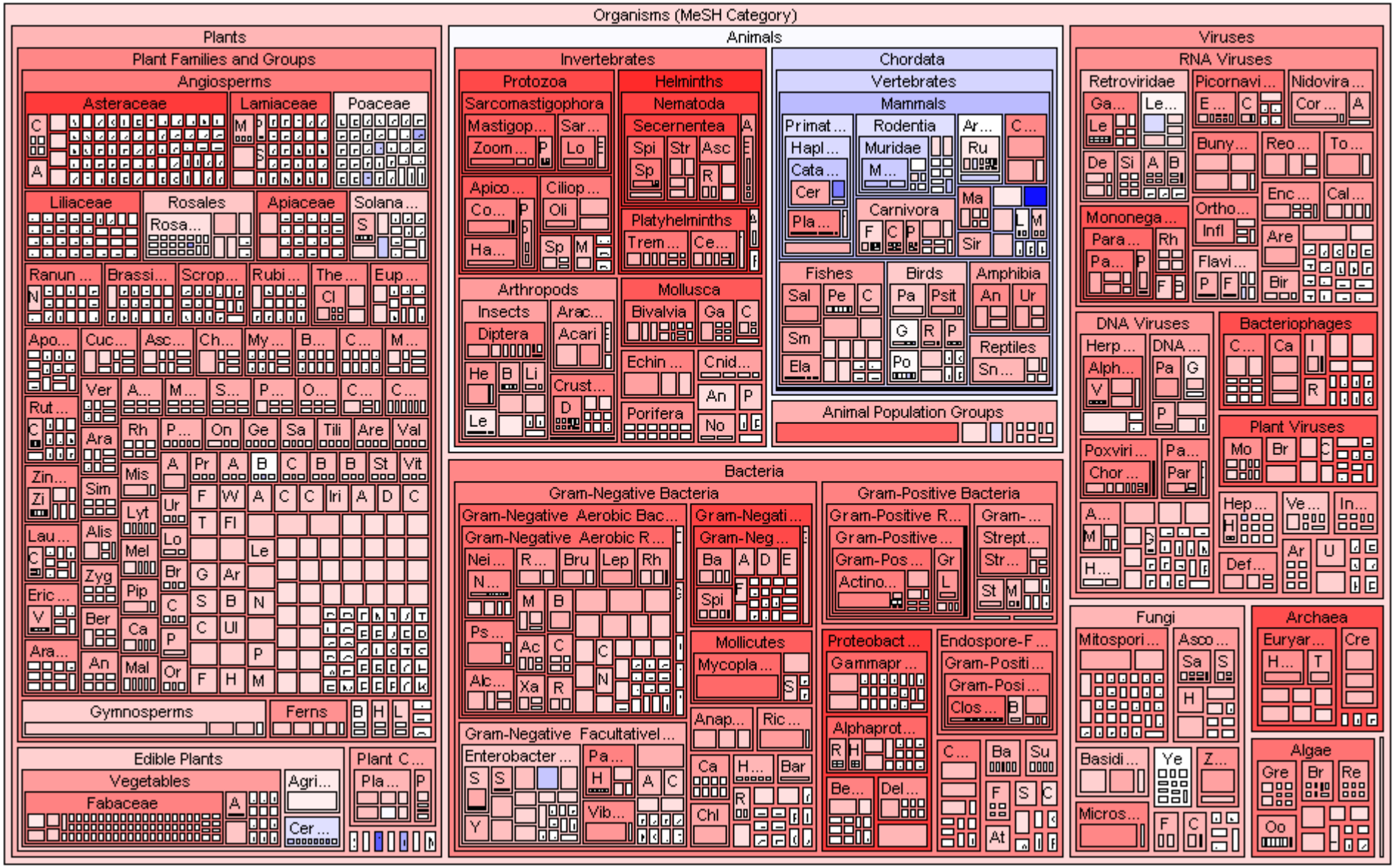
IC Diff


MeSH thesaurus using XML sourcefile date: Jul 10, 2006 1:56:24 AM
Chemicals and Drugs (MeSH Category)
Amino Acids, Peptides, and Proteins
Proteins
Peptides
Organic Chemicals
Heterocyclic Compounds
Enzymes and Coenzy...
Enzymes
Carbohydrates
Biological Fac...
Inorganic Chemicals
Chemical A...
Lipids
Diseases (MeSH Category)
Bacterial Infectio...
Neoplasms
Congenital, H...
Nervous Sy...
Digestive ...
Patholog ...
Disorder ...
Musculo...
Stomatog...
Virus Dis ...
Cardiovasc...
Skin and...
Eye Dise...
Hemic and...
Organisms (MeSH Category)
Plants
Plant Families and Gr...
Angiosperms
Bacteria
Gram-Neg...
Viruses
Fungi
Animals
Invertebrates
Chordata
Biological Sciences (MeSH Cat...
Health Occupations
Dentistry
Biologic...
Bioche...
Genetic...
Anatomy (MeSH Category)
Cells
Body R...
Nervous Sy...
Psychiatry and Psyc...
Anthropology, Educa...
Social Sciences
Analytical, Diagnostic a...
Diagnosis
Investigative T...
Therapeutics
Physical Sciences (...
Natural Sciences
Information...
Health Care (Me...
Geographic...

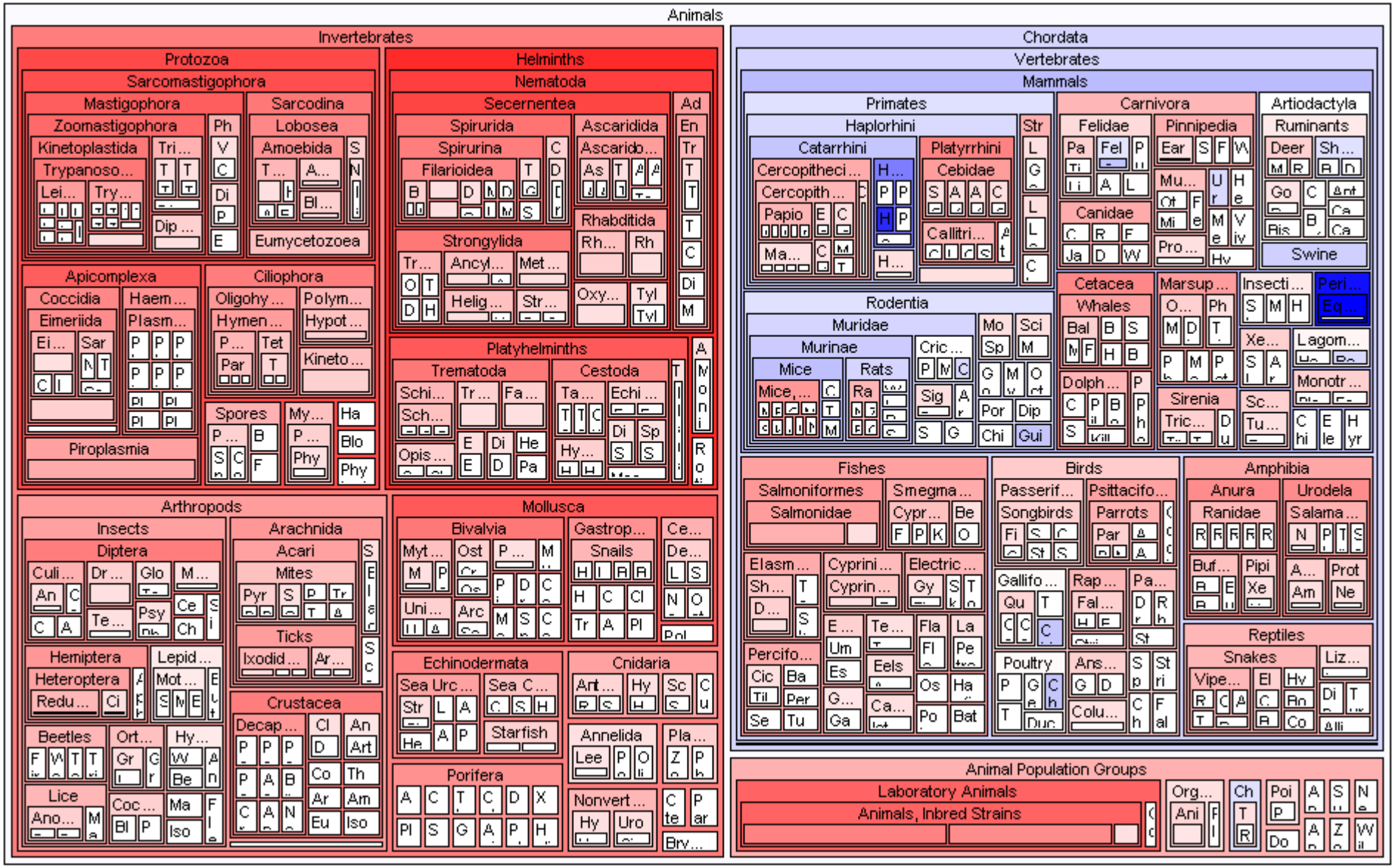
Bioscience


Biological Sciences (MeSH Category)
Health Occupations
Health Services Administration
Organization and Administr...
Professional Practice
Nursing
Nursi...
Patient Care Mana...
Delivery of Health...
Hospital Admini...
Hospital...
Perso...
Quality of Health C...
Health Care...
Evalu...
Quality Assu...
Dentistry
Prosthodontics
Dental Pr...
Oral...
Perio...
Ora...
De...
Den
Orthodontics
Endo...
Ope...
Dent...
Medicine
Medical Specialties
Physic...
Reha...
Psy...
Su...
Forens...
Pha...
Alli...
Vet
Env
Mor
Environment and Public Health
Public Health
Epidemiolo...
Demography
Pop...
Eth
Fa
Po
Re
Public Heal...
Commu...
Im...
Pe
Inf...
Environ...
E...
Air
W
F
Sanitation
Sani...
Ra...
Acc...
Epidemiolo...
Epidemiolo...
E...
Cl
Epide...
Epidemi...
Eff
C
R
Ag
B
Di...
Hea...
Hy...
Environment
Meteorol...
Atmosp...
W
C
Con...
Ho
S
Ener...
E...
Disas...
Biological Phenomena, Cell Phenomena, ...
Cell Physiology
Cell Cycle
Cell Div...
Cell N...
Cell...
H
G
C...
Blo
Cell
Biological Phe...
Microbiologic...
Drug...
Dr...
B...
Viral P...
Re...
Pi
Ev
Biol...
Immunity
Ce...
Tr
A...
Im
Co
N
Genetic Structures
Genome
Genome Components
Genes
Neo...
On...
Re...
G
D
Gen...
R
Viral G...
Viral...
Ba
O
F
P
Chromosomes
Mammalian...
Human C...
C
Chr...
Se...
Arti...
Pla
Ge
Circulatory and Respiratory...
Cardiovascular Physiology
Cardiovascu...
Blo...
He
M
Cardiov...
Hem...
B
Respiratory...
Respir...
T...
P...
Re
R
Blood Ph...
Blood P...
Reproductive and Urinary P...
Reproduction
Pregnancy
M...
O...
P
L
Se...
Cl
Me...
Est...
Fert...
In
Sexual B...
Musculoskeletal, Neural, an...
Musculoskeletal Ph...
Musculoskel...
Mu
Move...
M...
Ey
Ne...
Ocular Physiology
Ocular Phy...
Visual...
O...
Genetic Processes
Gene...
D
Mutag...
Geneti...
Br...
G
Genetic Phenomena
Variati...
Mutat...
Inher...
Rec...
Ge...
Pl...
Biochemical Phenomena, Metabol...
Metabolism
Biolog...
R
En...
Lipi
O
Peptid...
Prot...
Ph...
B
Ca
Biochemical Phenomena
N...
In
Bin
Se
Si...
Bo
Nutri...
Diet
Chi...
Physiological Pr...
Growth and D...
Morpho...
Emb...
E
Hu
Gr
Bo...
Ho
C
Digestive, Or...
Dig...
De...
Skin Physio...
Chemical a...
Dru

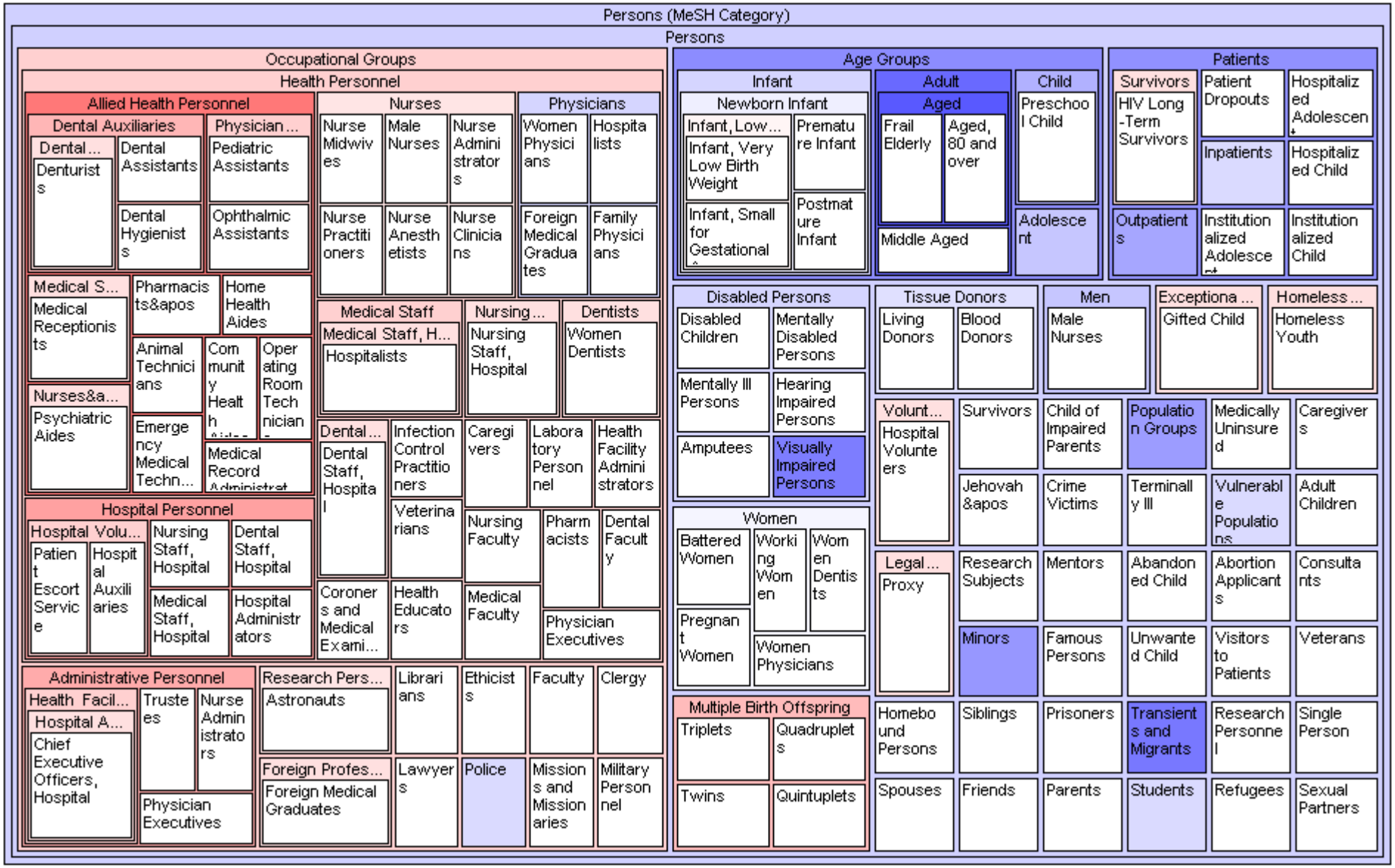
Organisms


Organisms (MeSH Category)
Plants
Plant Families and Groups
Angiosperms
Asteraceae
Lamiaceae
Poaceae
Liliaceae
Rosales
Apiaceae
Solana...
Ranun...
Brassi...
Scrop...
Rubi...
The...
Eup...
Gymnosperms
Ferns
Edible Plants
Vegetables
Fabaceae
Agri...
Cer...
Plant C...
Animals
Invertebrates
Protozoa
Helminths
Sarcomastigophora
Nematoda
Arthropods
Insects
Diptera
Arac...
Acari
Crust...
Mollusca
Bivalvia
Echin...
Cnid...
Porifera
Chordata
Vertebrates
Mammals
Primat...
Rodentia
Muridae
Carnivora
Fishes
Birds
Amphibia
Reptiles
Animal Population Groups
Bacteria
Gram-Negative Bacteria
Gram-Negative Aerobic Bac...
Gram-Negative Aerobic R...
Gram-Negati...
Gram-Neg...
Gram-Negative Facultativel...
Enterobacter...
Gram-Positive Bacteria
Gram-Positive R...
Gram-Positive...
Gram-Pos...
Actino...
Strept...
Mollicutes
Mycopla...
Proteobact...
Gammapr...
Alphaprot...
Endospore-F...
Gram-Positi...
Gram-Posi...
Clos...
Viruses
RNA Viruses
Retroviridae
Picornavi...
Nidovira...
Buny...
Reo...
Ortho...
Infl...
Flavi...
Mononega...
Para...
Enc...
Cal...
Are
Bir
DNA Viruses
Herp...
DNA...
Alph...
Poxviri...
Chor...
Bacteriophages
Plant Viruses
Mo
Br
Hep...
Ve...
In...
Def...
Fungi
Mitospori...
Asco...
Basidi...
Micros...
Archaea
Euryar...
Cre
Algae
Gre
Br
Re
Oo

Animals


Animals
Invertebrates
Protozoa
Sarcomastigophora
Mastigophora
Sarcodina
Zoomastigophora
Lobosea
Kinetoplastida
Amoebida
Eumycetozoea
Apicomplexa
Ciliophora
Coccidia
Eimeriida
Piroplasmia
Spores
Helminths
Nematoda
Secernentea
Spirurida
Ascaridida
Spirurina
Filarioidea
Strongylida
Rhabditida
Platyhelminths
Trematoda
Cestoda
Arthropods
Insects
Arachnida
Diptera
Acari
Mites
Ticks
Hemiptera
Heteroptera
Beetles
Lice
Crustacea
Mollusca
Bivalvia
Gastrop...
Snails
Echinodermata
Cnidaria
Starfish
Annelida
Porifera
Chordata
Vertebrates
Mammals
Primates
Haplorhini
Catarrhini
Platyrrhini
Cercopitheci...
Cebidae
Carnivora
Felidae
Pinnipedia
Canidae
Artiodactyla
Ruminants
Swine
Rodentia
Muridae
Murinae
Mice
Rats
Cetacea
Whales
Marsup...
Insecti...
Sirenia
Fishes
Salmoniformes
Salmonidae
Birds
Passerif...
Psittacifo...
Songbirds
Parrots
Amphibia
Anura
Urodela
Ranidae
Reptiles
Snakes
Poultry
Animal Population Groups
Laboratory Animals
Animals, Inbred Strains

Persons


Persons (MeSH Category)
Persons
Occupational Groups
Health Personnel
Allied Health Personnel
Dental Auxiliaries
Dental ...
Denturists
Dental Assistants
Dental Hygienists
Physician ...
Pediatric Assistants
Ophthalmic Assistants
Medical S...
Medical Receptionists
Pharmacists&apos
Home Health Aides
Animal Technicians
Community Health
Operating Room Technician
Nurses&a...
Psychiatric Aides
Emergency Medical Techn...
Medical Record Administrat
Hospital Personnel
Hospital Volu...
Patient Escort Service
Hospital Auxiliaries
Nursing Staff, Hospital
Dental Staff, Hospital
Medical Staff, Hospital
Hospital Administrators
Administrative Personnel
Health Facil...
Hospital A...
Chief Executive Officers, Hospital
Trustees
Nurse Administrators
Physician Executives
Nurses
Nurse Midwives
Male Nurses
Nurse Administrators
Nurse Practitioners
Nurse Anesthetists
Nurse Clinicians
Physicians
Women Physicians
Hospitalists
Foreign Medical Graduates
Family Physicians
Medical Staff
Medical Staff, H...
Hospitalists
Dental...
Dental Staff, Hospital
Infection Control Practitioners
Veterinarians
Coroners and Medical Exami...
Health Educators
Nursing ...
Nursing Staff, Hospital
Caregivers
Laboratory Personnel
Nursing Faculty
Medical Faculty
Dentists
Women Dentists
Health Facility Administrators
Pharmacists
Dental Faculty
Physician Executives
Research Pers...
Astronauts
Librarians
Ethicists
Faculty
Clergy
Foreign Profes...
Foreign Medical Graduates
Lawyers
Police
Missions and Missionaries
Military Personnel
Age Groups
Infant
Newborn Infant
Infant, Low...
Infant, Very Low Birth Weight
Infant, Small for Gestational
Premature Infant
Postmature Infant
Adult
Aged
Frail Elderly
Aged, 80 and over
Middle Aged
Child
Preschool Child
Adolescent
Disabled Persons
Disabled Children
Mentally Disabled Persons
Mentally Ill Persons
Hearing Impaired Persons
Amputees
Visually Impaired Persons
Women
Battered Women
Working Women
Women Dentists
Pregnant Women
Women Physicians
Multiple Birth Offspring
Triplets
Quadruplets
Twins
Quintuplets
Patients
Survivors
HIV Long-Term Survivors
Patient Dropouts
Hospitalized Adolescent
Inpatients
Hospitalized Child
Outpatients
Institutionalized Adolescent
Institutionalized Child
Tissue Donors
Living Donors
Blood Donors
Men
Male Nurses
Exceptiona ...
Gifted Child
Homeless ...
Homeless Youth
Volunt...
Hospital Volunteers
Survivors
Child of Impaired Parents
Population Groups
Medically Uninsured
Caregivers
Jehovah&apos
Crime Victims
Terminally Ill
Vulnerable Populations
Adult Children
Legal ...
Proxy
Research Subjects
Mentors
Abandoned Child
Abortion Applicants
Consultants
Minors
Famous Persons
Unwanted Child
Visitors to Patients
Veterans
Homebound Persons
Siblings
Prisoners
Transients and Migrants
Research Personnel
Single Person
Spouses
Friends
Parents
Students
Refugees
Sexual Partners