

Technical Report



**Distilling Subject Concepts
from OpenCyc**

Volume 3

Appendices

TR 08-07-16-B3

July 2008

Acknowledgements

The UMBEL project would like to thank Zitgist LLC for its generous donation of time and resources in programming and writing the documentation for this Technical Report.

UMBEL would also like to thank Cycorp for its support and preparation of a more current OWL version of the OpenCyc knowledge base. We would especially like to thank Larry Lefkowitz for his internal advocacy and answering many questions.

The Cyc Foundation, notably Mark Baltzegar, has been instrumental in helping to guide us through OpenCyc and to share with us many Foundation resources and projects currently in progress. The effort to date would not have been possible without this assistance.

- Michael K. Bergman, editor
- Frédéric Giasson, editor

UMBEL (Upper Mapping and Binding Exchange Layer) is a lightweight ontology structure for relating Web content and data to a standard set of subject concepts. Its purpose is to provide a fixed set of reference points in a global knowledge space. These subject concepts have defined relationships between them, and can act as binding or attachment points for any Web content or data.

*Connecting to the UMBEL structure thus provides **context** to Web data. In this manner, Web data can be linked, made interoperable, and more easily navigated and discovered. The project Web site is at <http://www.umbel.org>.*

UMBEL defines “subject concepts” as a distinct subset of the more broadly understood concept such as used in the SKOS RDFS controlled vocabulary or formal concept analysis or the very general concepts common to some upper ontologies. Subject concepts are a special kind of concept: ones that are concrete, subject-related and non-abstract. We further contrast these with named entities, which are the real things or instances in the world that are members of these subject concept classes. The UMBEL “backbone” is this set of reference subject concepts.





Table of Contents

REPORT SERIES OVERVIEW.....	1
APPENDIX A: FILES DOCUMENTATION.....	3
APPENDIX B: LISTING OF ABSTRACT CONCEPTS	4
APPENDIX C: NAMED ENTITY ASSIGNMENTS	11
Excluded from Consideration	11
SC Assignments	11
NE Assignments	11
APPENDIX D: OPENCYC SCORING CATEGORIES	12
1 - Subject Concepts (SC).....	12
2 - Internal Cyc	12
3 - Not useful UMBEL concepts	12
4 - Specific projects	12
5 - Verbs, predicates, Fn	12
6 - Software and programs	13
7 - Games	13
8 - Open	13
9 - Named Entities (NE).....	13
10 - Temporary	13
86 - Remove	13
99 - Unclassified	13
APPENDIX E: OPENCYC CATEGORIZATION ISSUES	14
OpenCyc OWL Mapping: The "Missing"	14
Excessive 'Cruft'	14
Treatment of NARTs.....	15
Ambiguous Areas	15
APPENDIX F: GENERAL PROCESSING TIPS.....	16
OpenCyc CSV Files.....	16
APPENDIX G: CYTOSCAPE USE TIPS.....	17
APPENDIX H: UMBEL WEB SERVICES.....	18

REPORT SERIES OVERVIEW

This document is one of three volumes in the Technical Report series, *Distilling Subject Concepts from OpenCyc*. These three volumes, in combination with the ontology documentation, complete the release package for UMBEL version 0.70, released July 16, 2008. This is the first release of the ontology.

This current three-volume series describes the selection and vetting of UMBEL's 20,000 subject concepts from OpenCyc.¹ These three volumes are:

- *Distilling Subject Concepts from OpenCyc, Vol. 1: Overview and Methodology*, **TR 08-07-16-B1**, the basic introduction and explanation of terminology and the distillation process
- *Distilling Subject Concepts from OpenCyc, Vol. 2: Files Documentation*, **TR 08-07-16-B2**, the listing and description of the various files accompanying this process, and
- *Distilling Subject Concepts from OpenCyc, Vol. 3: Appendices*, **TR 08-07-16-B3**, this volume, which provides supporting materials and detailed backup.

¹ <http://opencyc.org>



In addition, key specifications for the UMBEL ontology itself are documented in two volumes:

- *UMBEL Ontology, Vol. 1: Technical Documentation*, **TR 08-07-16-A1**, that overviews the ontology schema, vocabulary and use, and
- *UMBEL Ontology, Vol. 2: Subject Concepts and Named Entities Instantiation*, **TR 08-07-16-A2**, which is an explanation of the N3 files in the ontology distribution.



APPENDIX A: FILES DOCUMENTATION

See Volume 2 for complete files documentation.

APPENDIX B: LISTING OF ABSTRACT CONCEPTS

Here is the listing of the 739 abstract concepts within the UMBEL structure:

AbductiveInferenceProcess	AnimalByTaxonomicPhylumWithSamples-Biology-Topic
AbnormalSystemCondition	AnimalCapabilityType
Abnormal-Unusual	AnimalPhysiology-Physiology-Topic
AbstractIndividual	AnimalsByGrossForm-Biology-Topic
AbstractInformationStructure	AnimalsByMiscellaneousCriteria-Biology-Topic
AcademicCycles-Time-Topic	AnimalSensoryQuantity
AcademicDepartmentTypeBySubject	AnimalShapedThing
AcademyAwardCategory	AnimalSuppliesMarketCategory
AccomplishmentType	AntiqueFurnitureMarketCategory
AccountSystemTypeByImplementationMeans	ApplicationSpecific-Topic
AccountTypeByStatus	ArchitecturalTypeByStyle
Action	Area
Action-Device-Topic	Area_with_historical_buildings
ActionOnObject	AreaOfOperations
Actions-Topic	ArithmeticFunctions-Math-Topic
ActionTypeByLegality	Artifact
Activism	Artifact-AnimalCreated
ActorActressTypeByPerformanceType	Artifact-Generic
ActorTypeByMedium	Artifact-HumanCreated
ActTypeCommonlyConsideredCriminal	Artifact-Intangible
Addresses-Geography-Topic	Artifact-NonAgentive
AerodromeConcept	Artifacts-Topic
Agent-PartiallyTangible	ArtifactType
Agents-Topic	ArtifactTypeByFunction
Agreement-Topic	ArtifactTypeByHistory
AilmentConditionTypeBySyndrome	ArtificialMaterial
Ailments-Topic	Artist-PerformerTypeByPerformanceType
AircraftTypeByBrand	ArtistTypeByArtForm
AlgebraicNumber	ArtSuppliesMarketCategory
AllFemaleSchool	AspatialInformationStore
Allotment-Collective	Assignment-Obligation
AlternateRoute	Assignment-Specification
Alternative	AthleteTypeBySport
AlternativeMedicineEventTypeByModality	AthleticPhysicalBuild
AmateurSportsTeamTypeBySponsor	AtLeastPartiallyMentalEvent
AminoAcidTypeByChemicalSpecies	Attitude-Topic
AnatomicalVessel	AuthorizedAgreement
AngularAccelerationRate	AuthorTypeByFormat
AngularDistance	AuthorTypeBySubject
AngularDistanceBetweenTwoVectors	AutoFilterMarketCategory
AnimalBLOTTypeByPosition	AutoMechanicShopMarketCategory
AnimalBodyPartSystems-Physiology-Topic	AutomobileTypeByBodyStyle
AnimalBodyPartType	AutomobileTypeByBrand
AnimalBodyRegion	AutomobileTypeByConventionalSizeClassification
AnimalBodyRegionType	AutomobileTypeByModel
AnimalBodyShapeType	AutoSoundMarketCategory
AnimalBySex-Biology-Topic	AutoTireMarketCategory
AnimalByStageOfDevelopment-Biology-Topic	

AxisOfRotation-Earth
 AxisymmetricObject
 BachelorOfArtsLevel
 BachelorOfScienceLevel
 BakeryMarketCategory
 BallTypeByGame
 BaseballPlayerTypeByPosition
 BaseWordForm
 BasicOfficeConnector
 Basics-Fundamental-Topic
 BathroomFurnishingsMarketCategory
 BedroomFurnitureMarketCategory
 BedsAndMattressesMarketCategory
 BedTypeBySize
 Beliefs-Attitude-Topic
 BeliefSystemType
 BicycleMarketCategory
 BidTypeByStatus
 BiologicalClass
 BiologicalDevelopmentProcess
 BiologicalEvent
 BiologicalKinRelationships-Biology-Topic
 BiologicalLivingObject
 BiologicalProcess
 BiologicalProcessType
 BiologicalSpecies
 BiologicalSubclass
 BiologicalSubdivision
 BiologicalSubfamily
 BiologicalSubkingdom
 BiologicalSuborder
 BiologicalSubphylum
 BiologicalSubspecies
 BiologicalSuperfamily
 BiologicalTaxon
 Biology-Topic
 BirthControlMarketCategory
 BloodTypeByABO
 BloodTypeByABOAndRhFactor
 BloodTypeByRhFactor
 BodyPositionAndMovement-Physiology-Topic
 BoneSubPart
 BooksAndMagazinesMarketCategory
 Borderline
 Border-StripNearTheBoundaryOfAnObject
 BoundaryLine
 Boundary-Underspecified
 BoundedDirectedAcyclicGraph
 BoundedDirectedPathSystem
 BrandNamePharmacyProductType
 BridgeTypeBySupportType
 BuildingConcept
 Business-Fortune500
 BusinessRelatedThing
 BusinessTypeByActivityType
 BusinessTypeByManufacturingIndicator
 BuyingAnObject
 Bypass-Path
 CalendarSeasonType
 CannedGoodsMarketCategory
 CarbamateCompoundType
 CartographicFeatureType
 CasualClothingMarketCategory
 CatTypeByBreed
 Causality-Attitude-Topic
 CausingAnotherObjectsTranslationalMotion
 CelestialEvents-Time-Topic
 CelestialRegion
 CellularProcessType
 ChangeOfStatusEvent
 Check-TenderObject
 ChemicalAgentType
 ChemicalCompoundsMolecules-Chemistry-Topic
 ChemicalCompoundTypeByChemicalSpecies
 ChemicalObjectTypeByChemicalSpecies
 ChemicalReactions-Chemistry-Topic
 ChemicalReactionTypeByEquationFormula
 ChemicalSpeciesType
 Chemistry-Topic
 ChildrensFurnitureMarketCategory
 CircularRegion
 ClassicalMusicalCompositionType
 ClassicalMusicalCompositionTypeByMusicalForm
 ClimateCycles-Weather-Topic
 ClimaticSeasons-Time-Topic
 ClimaticSeasons-Weather-Topic
 CloseContactTransmittedInfection
 ClosedFacility
 ClosedLinearObject
 ClosedObject
 ClosedPortal
 ClosedSurfaceRegion
 ClothingMarketCategory
 Clothing-Topic
 CloudlikeObject
 Collapsible
 CollapsibleDevice
 Collapsing-Generic
 Collection
 CollectionType
 CollectionTypeType
 ColoredThing
 ColorlessThing
 CommercialActivities-Financial-Topic
 CommonSense
 Communicating
 Communication-Topic
 Comparisons-Quantity-Topic

Competition-Social-Topic
 Composition-Topic
 CompressedSubstance
 ComputationalObject
 ComputerFileTypeByFormat
 ComputerHardwareMarketCategory
 ComputerMarketCategory
 ComputerSoftwareMarketCategory
 Computers-Topic
 ConceptualAbstraction
 ConceptualWorkSeries
 Condensing
 Configuration
 ConfigurationOfComplexObject
 ConformingToASpecification
 ConnectednessPreservingProjection
 ConnectedPathSystem
 ConnectedPhysicalPathSystem
 Connection-Physical
 Connections-Spatial-Topic
 ConnectionsToTheComputationalSubstrate-
 Fundamental-Topic
 Connector
 ConstructedHumanLanguage
 ConstructionArtifact
 Construction-Topic
 ConsumableProduct
 ConsumingFood-Food-Topic
 ContainerArtifact
 Container-Underspecified
 ContainingSomething
 Content
 Context-Topic
 ContinuousTimeInterval
 ControllingSomething
 ControversialAction
 ConventionalClassificationType
 CoordinatedBodyMovement
 CordlikeObject
 CosmeticsMarketCategory
 Covering-Object
 CreatingAnArtifact
 Criterion
 Cross-ConstructionArtifact
 CruciformObject
 Crystalline
 CulturalThing
 CurrentSource
 CurvaceousPhysicalBuild
 CyclicPolygon
 CycMetaphysics-Topic
 DateFunctions-Time-Topic
 DensityOfPopulation
 Device-FunctionallyDefective
 Device-Locked
 Device-OneTimeUse
 Device-Overheated
 Device-SelfControlled
 Device-SingleUser
 Devices-Topic
 Device-UserControlled
 Device-UserPowered
 DeviceWithNoMovingParts
 DevisedPracticeOrWork
 DevisedStructuredActivity
 DiffuseIllumination
 DiningRoomFurnitureMarketCategory
 DirectedMultigraph
 DirectedPath-Generic
 DirectedTranslation
 DisplayingSomething
 DisposableProduct
 Distance
 DividedPath
 DurativeEventType
 DynamicIndexedInfoSource
 EconomicSector-Localized
 EconomicSupport
 Education-Social-Topic
 ElectricalConductivity-Material-Topic
 ElectricalSupplyMarketCategory
 ElementsAndAtoms-Chemistry-Topic
 Emotion-Topic
 EnduringThing-Localized
 Engagement-Military
 Ensemble
 EventOrRoleConcept
 Evidence-Indication
 Exchange_
 ExistingObjectType
 Explanation
 ExternalAnatomicalPart
 ExtrudedBodyPart
 FailureOutcome
 FatPhysicalBuild
 FeminineHygieneMarketCategory
 FeminineObject
 Fermenting
 FictionalThing
 FieldFormats-Db-Topic
 FieldsOfKnowledge-Social-Topic
 FileManagement-Topic
 FinancialAccountTenderObject
 Financial-Topic
 FirstAidMarketCategory
 FiscalCycles-Time-Topic
 FlatObject
 FlatPhysicalSurface

FlatRegion	HostedSocialGathering
FlatSurfaceRegion	HouseholdLinensMarketCategory
FluidFiltrationProcess	HousewaresMarketCategory
FluidTangibleThing	HumanAccessibleIBO
FollowingAStandardProcedure	HumanAccessibleIBT
FoodAndDrink-Topic	HumanActivities-Topic
FoodProductTypeByBrand	HumanActivity
Foodstuffs-Food-Topic	HumanBeings-Topic
Food-Topic	HumanlyOccupiedSpatialObject
FootwearMarketCategory	HumanOccupationConstruct
ForceQuantity	HumanScaleObject
ForceVector	HumanSelfControlStandard
FormalityOfSpeech	HumanShelterConstruction
FormalProductType	HungerAndMalnutrition-Food-Topic
Form-StandardizedIBO	IBTGeneration
FreeThing	ImitationOfIndividual
FrictionVector	Improvement-Transformation
FrozenDessertMarketCategory	Inaction
FrozenFoodMarketCategory	InanimateObject
FunctionalAlgorithm	InanimateObject-Natural
FunctionalRelation	InanimateObject-NonNatural
Function-Denotational	IncreaseAction
Functions-Fundamental-Topic	IndependentThing
Fundamental-Topic	Individual
FurnitureMarketCategory	IndividualAgent
FutonsMarketCategory	IndividualDenotingFunction
GamesAndPuzzlesMarketCategory	Indulgence
GardeningSuppliesMarketCategory	IndustrialConcept
GeneralAnimalAnatomy-Physiology-Topic	Industry-Localized
GeneralAutoMarketCategory	InfiniteSpatialThing
GeneralMedicine-Topic	InformationBearingObject
GenericPharmacyProductType	InformationBearingPerformance
GenericPharmacyProductType-NameOnly	InformationBearingThing
GeographicalAgent	InformationBearingWavePropagation
GeographicalDirection	InformationDesignation
GeographicalDirection-General	InfrastructureIndustry
Geography-Topic	InfrastructureProduct
Granular	Ingesting
GroceryMarketCategory	InorganicMaterial
Group	InstrumentalArtifact
HairCareMarketCategory	IntangibleExistingThing
HardcopyInformationBearingObject	IntangibleIndividual
HardwareMarketCategory	IntelligentAgent
HealthAndBeautyMarketCategory	IntelligentAgentActivity
Healthcare-Topic	InterAgentSupportQuantity
HealthFoodMarketCategory	Interests-Social-Topic
HelpingAnAgent	InternetSoftwareMarketCategory
HexalateralObject	InvoluntaryBodyMovement
HistoricTemporalThing	IrrationalNumber
Holdings	JunctionOfPaths
Holidays-Time-Topic	KitchenFurnishingsMarketCategory
HomeElectronicsMarketCategory	KnickKnackMarketCategory
HomogeneousStructure	Knowledge-Attitude-Topic
HosieryMarketCategory	LandlockedGeopoliticalEntity

LargeApplianceMarketCategory
 LaundryEquipmentMarketCategory
 LaundrySuppliesMarketCategory
 LawnAndGardenMarketCategory
 Lexicon-Topic
 LifeStages-Social-Topic
 LifeThreateningCondition
 LightingMarketCategory
 LinearObject
 Line-Directed
 LineOfObjects
 LineSegment
 LinguisticConceptExtraction
 LinguisticObject
 LinguisticQuantity
 Linguistic-Topic
 LiquidSolution
 Liquid-StateOfMatter
 LiquidTangibleArtifact
 LiquidTangibleThing
 LogicAndMath-Topic
 LookingForAThingOrTypeOfThing
 LossOfUserRights-PossTransfer-Topic
 LuggageMarketCategory
 MachineProtocol
 MailableObject
 MaintainingSomething
 MakingSomething
 ManMadeThing
 MarketTypeByProductSold
 MasculineObject
 MassOrSpatialQuantity
 Material-Topic
 MathematicalFunctions-Fundamental-Topic
 MathematicalObject
 MathematicalOrComputationalThing
 MathematicalStatement
 MathematicalThing
 Math-Topic
 MeatCounterMarketCategory
 MedicalCareProviders-GeneralMedicine-Topic
 MedicalMarketCategory
 Medical-Topic
 Medicine-Other-Topic
 MensFormalWearMarketCategory
 MereotopologyConcept
 Metal-Chemistry-Topic
 MicroscopicScaleObject
 MiddleClass
 MilitaryOrganizationConcept
 MilitaryThing-SpatiallyLocalized
 MiscellaneousOrganizationConstants-Organization-Topic
 Mixture
 MobilityCorridor
 MonetaryRate
 MoneyAndCommerce-Topic
 MoneyAndIncome-Financial-Topic
 MoneyAndPrice-PossTransfer-Topic
 MotionPictureAwardCategory
 Movement-Topic
 MovieTypeByGenre
 MultiIndividualAgent
 MultiIndividualAgent-Intelligent
 MultiPurposeDevice
 MusicalInstrumentMarketCategory
 MythologicalThing
 NailCareMarketCategory
 Narrative
 NaturalLanguageProcessing-Topic
 NaturalResourceType
 NaturalThing
 NaturalWorld-Topic
 NonAspectualQuantity
 NonFluidlikeTangibleThing
 NonintrusiveJunction
 NonNaturalThing
 NonPersonAnimal
 NonPoweredDevice
 NonPublishedText
 NonZeroNumber
 NormativeSpecification
 NuclearRelatedThing
 ObjectSpecification
 Object-SupportingFurniture
 ObjectType
 Obligation
 ObtainingAndStoringKnowledge
 OfficeProductMarketCategory
 OfficialDocument
 Opaque
 OpenArea
 OpenSpaceContent
 Opinion-Attitude-Topic
 OpticalMarketCategory
 OrganismByHabitat-Biology-Topic
 OrganismBySourceOfNutrients-Biology-Topic
 OrganismByTaxonomicKingdom-Biology-Topic
 OrganismPart
 OrganismParts-Biology-Topic
 OrganismVarietyBySize
 Organization-AnimalActivity
 OrganizationBySector
 OrganizationOfPeopleOnly
 Organization-Topic
 OrganizationType
 OrganizationTypeByActivityType
 OutdoorAccessoryMarketCategory

OutdoorFurnitureMarketCategory
 PackagedCerealMarketCategory
 PackagingContainerProduct
 PartiallyIntangible
 PartiallyIntangibleIndividual
 PartiallyTangible
 PartiallyTangibleTypeByPhysicalFeature
 Particle
 Partition-ConstructionArtifact
 Parts-Topic
 PathArtifact
 PathArtifactSystem
 Path-Customary
 PathForWheeledVehicles
 Path-Generic
 Path-InfoTransfer
 Path-Spatial
 PeopleRelatedToOrganizations-Organization-Topic
 PerCapitaRatioQuantity
 Perception-Topic
 Permission
 PersonalProduct
 PersonalQuantity
 PersonTypeByPersonalityTrait
 PersonWithOccupation
 PetProductMarketCategory
 PhysicalDatabases-Db-Topic
 PhysicalField
 PhysicalFilter
 PhysicalImpactBuffer
 PhysicalLayerProtocol
 PhysicalPartOfObject
 PhysicalQuantity
 PhysicalSituation
 PhysicalStuffAndObjects-Topic
 PhysicalTransformationProcess
 Physics-Topic
 Physiology-Topic
 PlanarPointSet
 PlantPart
 PlantsByBotanicalTaxonomy-Biology-Topic
 PlantsByGrossForm-Biology-Topic
 PlantsBySeasonalChange-Biology-Topic
 PluralityOfLikeSubevents
 Policy
 PolyAtomicChemicalObject
 PolygonSet
 PopulationClassificationTerms-Social-Topic
 PortableObject
 Portal
 PossTransfer-Topic
 PreOwnedForSaleObject
 PrescriptionDrugMarketCategory
 Probability-Topic
 ProcessType
 ProduceMarketCategory
 ProductsAndMarketing-Financial-Topic
 ProductType
 ProductTypeByBrand
 ProductTypeByBrandVersion
 ProductTypeByMarketCategory
 Professional
 Professional-Adult
 ProfessionalCulture
 Professional-Topic
 PropositionalAttitudes-Topic
 PropositionalConceptualWork
 PropositionalInformationThing
 Proposition-ConceptuallyTrue
 Proposition-True
 ProtectionSituation
 Purity-Chemistry-Topic
 Purity-Composition-Topic
 PurposefulAction
 PurposefulPhysicalAction
 Quantity-Topic
 Question-IBT
 RadiallySymmetricObject
 RaisingLivingThings
 Rash-Object
 RasterArray
 ReclinerChairsMarketCategory
 RelationalStructures-Math-Topic
 Relations-Fundamental-Topic
 RelationshipsAmongOrganizations-Organization-Topic
 Relationships-Social-Topic
 Residence-Social-Topic
 Resources
 RestaurantOrg-WithoutWaitstaff
 RestaurantOrg-WithWaitstaff
 RevengeAction
 RowOfObjects
 ScalarDenotingFunction
 ScalarOrVectorInterval
 ScalarQuantity
 SchoolSuppliesMarketCategory
 ScienceAndNature-Topic
 SemiSolidTangibleThing
 Sensors-Perception-Topic
 SerialNumber-String
 SetOrCollection
 SetOrCollectionType
 SetTheory
 SewingSuppliesMarketCategory
 ShavingSuppliesMarketCategory
 SheetOfSomeStuff
 ShortPhysicalBuild

SinglePurposeDevice
 SingleSiteOrganization
 Situation
 SituationalObjectType
 Situation-Localized
 SkinCareMarketCategory
 SmallApplianceMarketCategory
 SnackFoodMarketCategory
 SocialBeing
 SocialClass-Social-Topic
 SocialOccurrence
 Social-Topic
 SoftwareObject-Individual
 SolidAnimalMatter
 Solid-StateOfMatter
 SolidTangibleArtifact
 Solubility-Material-Topic
 SomethingExisting
 SomethingToWearPart
 SoundInformationBearingThing
 SpatialThing-Localized
 Spatial-Topic
 SpecificAnimalBodyParts-Physiology-Topic
 SpecificPhysiologicalProcesses-Physiology-Topic
 SportingGoodsMarketCategory
 StandardAutoMarketCategory
 StandardAutoServiceMarketCategory
 States-Device-Topic
 StationeryMarketCategory
 Strategy
 StrictlyMentalEvent
 StringObject
 StripNearTheBoundaryOfAnObject
 StructuredKnowledgeSource
 StudentByEducationalOrganizationLevel
 StuffType
 Sub-AtomicParticles-Chemistry-Topic
 Surface
 SurfaceForceVector
 SurfaceGeometryConcept
 SurfaceRegion-Infinite
 SymbolicObject
 SymbolicThing
 System
 SystemOfDevices
 TacticalArea
 TakingCareOfSomething
 TallPhysicalBuild
 Technology-Artifact
 TemporallyContinuousThing
 TemporallyExtendedThing
 TemporalObjectType
 TemporalStuffType
 TemporalThing
 TenderObject
 TerrainConcept
 Terrorism-Topic
 TheCalendar-Time-Topic
 ThePhysicalUniverse
 ThermalConductivity-Material-Topic
 Thing
 Thinking
 Thinking-NonPurposeful
 TimeInterval
 TimePoints-Time-Topic
 Time-Quantity
 Time-Topic
 TitlesAndNames-Social-Topic
 TobaccoSuppliesMarketCategory
 ToothCareMarketCategory
 ToysMarketCategory
 TranscendentalNumber
 TransportationContainerProduct
 Transportation-Topic
 TravelProductsMarketCategory
 TriggerableDevice
 TupleOfIntervals
 UniqueID
 UnitedStates-Geography-Topic
 UnitOfMeasureConcept
 UnitVectorInterval
 UniversalConstant
 UnversionedSoftwareObject
 UpperClass
 UsedFurnitureMarketCategory
 UtilityFurnitureMarketCategory
 VectorField
 VectorInterval
 Vector-Precise
 VectorQuantity
 Vectors-Math-Topic
 VertebratesByTaxonomicClass-Biology-Topic
 VideoRecordedObject
 VisibleFeatureType
 VisualInformationBearingThing
 VisualInformationGeneratingAct
 WaterbedMarketCategory
 WaterBodyConcept
 WaterSurfacePath
 WaterSurfaceTransportation
 Wave-Topic
 WeatherAttributes-Weather-Topic
 WeatherEvents-Weather-Topic
 WeatherObjects-Weather-Topic
 Weather-Topic
 WheeledTransportationDevice
 WorkingProcess

APPENDIX C: NAMED ENTITY ASSIGNMENTS

This appendix records specific instances of assignment splits between named entities (NEs) and subject concepts (SCs).

Excluded from Consideration

- Generally, NO action is an object
- Moods and feelings and attitudes
- Similarly, the qualitative "attributes" of things such as color, shape, aspects, quantity (excepts as concepts in their own rights)
- Grammar and NL-related
- "Level of"

SC Assignments

These ambiguous entries were assigned to Subject Concepts:

- Protocols
- Geo eras
- Large geo areas (countries and larger, such as regions and continents)
- Car by manufacturer, but not model
- Beetles and orchids
- Languages and language families
- Religions
- Chemical compounds (still open)

NE Assignments

These ambiguous entries were assigned to Named Entities:

- Dealerships
- Dictionaries
- States, counties and cities; all other named places
- Genus and below
- Leafs and seeds
- Most ranks
- Constellations.

APPENDIX D: OPENCYC SCORING CATEGORIES

These are the marked types assigned to the deletion categories for OpenCyc collections (classes) and individuals. Notes for determining deletion are provided for each.

1 - Subject Concepts (SC)

These entries are OK and are the object of the entire subject concept processing workflow. In early list versions, these are also known as "accepted", when they are intended to eventually be transferred to 'clean' or 'final' lists at the conclusion of review.

2 - Internal Cyc

These are internal Cyc concepts or entities. 'Cyc' text find, plus certain internal constructs like the TextLearner, etc., are used as a first cut FIND in text screening. Some of the FIND text strings of use in this category are Template, Mt (microtheory), argXXX, genXXX, interArg, fACT, FACT, etc.

3 - Not useful UMBEL concepts

Generally, these entries are either too specific or too abstract for UMBEL purposes. Examples of entries that should be deleted are:

- Abstract entries - such as Artifact, Thing, Collection, Set, State, Belief, Some__, etc.
- Modifying concepts - such as Partial, Moderate, Heavy, Solid, Fixed, Strong, Very, Highly, etc. These are almost like nuanced or adjectival concepts
- Certain 'XXX_of' concepts - such as Group_of, Map_of, Resident_of, Border_, Center_the_glyph_, etc.
- Clearly specific concepts not related to Subject_Concepts_v_Named_Entities
- FoodServiceEstablishment - these are nearly all local
- Most initial lowercase URL names

This deletion category overlaps potentially with the next one.

4 - Specific projects

There are a tremendous amount of OpenCyc entries obviously related to past or current projects in military, intelligence, computer security, Pfizer, porno, etc. This often displays through specific microtheories or named entity listings (such as weapon types) obviously imbalanced with the level of specificity for entities in other categories.

Some of the text strings useful to FINDing these projects are:

- BBN, Brandeis, BUTLER, COA, CADRE, ChatterbotStateMachine, GlassBox, COLT, GNDB, CSC, VPQ, NCR, LF, EDB

5 - Verbs, predicates, Fn

Predicates and functions (Fn) are labeled as such and relatively easy to flag through a FIND text and mark. Verbs are generally ones ending in "ing" such as "editing" or "thinking"; these are clearly not noun or noun-phrased based "subject concepts" or "named entities".



In marking these for deletion, care should be taken for "ing" words that also substitute or work as concepts, such as "writing" or "shipping".

6 - Software and programs

Though perhaps strictly this category could be called a #4 above, there are enough of them to warrant a separate flagging. While broadly generic software or hardware should be included in the UMBEL listing, there are many specific entries that are obviously imbalanced with entities in other categories.

A text FIND can be applied to items such as TheProgram, TheSoftwarePackage, TheCGIScript, TheSoftwarePatch, TheEnvironmentVariable, WindowsService, Mod or ApacheModule, ApacheDirective or TheLibrary for an initial screening; other deletion candidates appear via manual inspection.

7 - Games

Though perhaps strictly this category could be called a #4 above, there are enough specific games to warrant a separate flagging. While broadly generic games should be included in the UMBEL listing, there are many specific entries that are obviously imbalanced with entities in other categories.

A text FIND can be applied to items such as TheGame and Game for an initial screening; other deletion candidates appear via manual inspection.

8 - Open

This category is reserved for latter use.

9 - Named Entities (NE)

This is the last review step in the pipeline, when "atomic" items are assigned to the Named Entity category. In early versions of the lists in the workflow, this assignment may be a temporary one meaning "re-check", or that the temporary assignment should be re-checked later.

10 - Temporary

This is a temporary assignment category used during review processing; it should never appear in a final deletion listing.

86 - Remove

This deletion number is used for malformed or bad URLs in the knowledge base.

99 - Unclassified

While these items are appropriate for deletion, it is not clear to which category above they should be assigned.

APPENDIX E: OPENCYC CATEGORIZATION ISSUES

Here are some issues with respect to using the OpenCyc knowledge base as originally encountered as a source of canonical "subject concepts" (SCs).

Nearly all issues identified in the early phases of reviewing OpenCyc have either been completely addressed, or are in the process of being so, by Cycorp itself in the scope of the OpenCyc knowledge base or the procedures used to create its OWL versions. Subsequent versions of OpenCyc should have very few residual concerns.

OpenCyc OWL Mapping: The "Missing"

Of the more than 173,000 OpenCyc entities, about one-third (or about 61,000) are "missing". We define "missing" as an OpenCyc entry that is a subject of a triple in the OWL version of the KB, but that is not an instance of the class *cyc:Individual* (without inferring this fact) nor an instance of the class *rdfs:Class*.

The inspection of these items suggests that most are Individuals in nature. This is due by the fact that we can infer, based on the OWL-Full description of OpenCyc, that they are effectively instances of the class *cyc:Individual*. However this fact was not apparent when we first come back these "missing" subject concepts.

Excessive 'Cruft'

This issue truly fits within the conventional definition of cruft, not in the sense of "poor quality", but in the sense of leftover, accumulated stuff from prior efforts. The general sense of working with Cyc is working with a venerable and well-constructed object (say, the Eiffel Tower), but one that has had too much dirt and accumulated rust and is therefore in need of a thorough steam cleaning. This cleaning could be superficial, or substantial and detailed. If the latter, there are prospects to bring actual shine and clarity to the solid ontology infrastructure at the core of the KB.

The perception of 'cruft' increases when OpenCyc is viewed through the lens of being an 'open-world' reference structure for binding and integrating external datasets. Under this perspective, the challenge is accessibility and understandability, not inferencing capabilities and use as a self-contained knowledge base.

To be sure, those aspects of OpenCyc that make it a complete knowledge base are an essential part of its attraction. To be able to take an external concept, and draw inferences upon it and utilize the advanced capabilities of OpenCyc or its more capable progenitors ResearchCyc or Cyc is the prime rationale for looking to OpenCyc as the canonical guiding framework for a reference subject concept backbone.

The challenge for OpenCyc becoming the go-to reference structure for diverse content must be met through a layered approach, or at least one of successive and progressively more complete access. Initially, understanding and accessing OpenCyc should be simple and direct. Once accessed, simple relationships can be shown and tested. Then, over time with use and familiarity, greater richness and inferencing capabilities may be brought to bear.

OpenCyc is daunting taken in one gulp. Moreover, much of what is contained in the system is an artifact of more than two decades of use and with respect to multiple and varied internal capabilities (also much of which is unlikely to be used or of interest to even veteran users).

The need is to peel back the layers of the onion and only expose as much as current needs and uses require. That is the purpose of the current analysis, rather than a broad criticism of the KB itself.

Fundamentally, the premise is that broad use of the KB requires more variety in access points and accessibility. Specific focus on this challenge may be perhaps the most critical aspect of how broadly Cyc becomes adapted by the linked data and semantic Web communities.

Treatment of NARTs

A NART is a reified NAT, and are implemented with data structures that have indexing allowing all uses of the NART to be retrieved.

A non-atomic term (NAT) is a way of specifying a term as a function of some other term(s). Every NAT is composed of a function and one or more arguments to that function.

Consider, for example, the function `#$FruitFn`, which takes as an argument a type of plant and returns the collection of the fruits of that type of plant. This function can be used to build the following NATs:

```
(#$FruitFn #$AppleTree)
($FruitFn #$PearTree)
($FruitFn #$WatermelonPlant)
. . . .
```

Note that there may or may not be a named Cyc constant corresponding to the collection of apples (that is, a constant called `#$Apple`). The NAT (`#$FruitFn #$AppleTree`) provides a way of talking about this collection even if the corresponding constant does not exist. NATs can be used anywhere a constant can be used.

In general, all NARTs and entries with an Fn designation were removed from SC consideration. NARTs and Fns should probably be reviewed for inadvertent exclusions.

Ambiguous Areas

- Are 'Templates' internal Cyc (#2) or #3 or even (in some cases) #4?
- Are 'Faults' internal to Cyc (#2) or otherwise?
- Are 'Frames' internal to Cyc (#2) or otherwise?

APPENDIX F: GENERAL PROCESSING TIPS

OpenCyc CSV Files

Note the # number references below apply to the Appendix C OpenCyc Scoring Categories.

- Using MS Excel or scalc from OpenOffice is helpful during manual screening. Here are a few processing and review tips:
 1. Keep files in CSV format (if XLS during review, that is OK, but convert back when completed for consistency and smaller file size)
 2. First sort any candidate review file alphabetically by URL
 3. Add an initial column to the left for score/mark assignments; add an additional left column for temporary processing via functions
 4. Use the VLOOKUP function for comparing sorted entries from two files. Make sure you give your reference block a name (for easier formula duplication and make sure the URLs are in its first column; the URLs provide the best comparison basis). Also, make sure the third argument in the function is FALSE to ensure exact compares
 5. The text FIND function is also highly useful for screening exact (case-sensitive) string inspections
 6. When functions are executed with results in the leftmost column, copy the entire column and then past back as 'values' prior to sorting
 7. Take care to ensure sorts were accurate and included all relevant columns
 8. Re-save your file frequently during review (once a sort is screwed up, it is impossible to re-do and you must return to an earlier saved version).
- In general, "out-of-scope" entries (#4, #6 and #7) are provided precedence for removal. Then, internal Cyc (#2) and predicates (#5) are assigned. Thus, a predicate in an external project entry would be assigned a #4 v #5.
- Try to keep notes regarding decision rules (see above) or ambiguous areas or open questions (see below) when manually inspecting a new list.

APPENDIX G: CYTOSCAPE USE TIPS

Our initial use of the tool suggests some use tips:

- For UMBEL, a system with 2 GB of RAM appears to be the minimal for creating new layouts (smaller amounts of RAM can be used for viewing the distribution packages). If your system has more than 512 MB of RAM available, you should up the start-up configuration for Cytoscape. See xxx for instructions
- Try Cytoscape with the yFiles layouts; quicker to perform, and interesting results
- Try the Organic yFile layout as one of the first
- Try the search feature
- Check the manual for examples of layouts
- Holding the right mouse button down when in the main screen; moving the cursor from the center outward causes zoom in, from the exterior inward, to zoom out
- Moving and panning nodes can be done in real time without issues
- The “edge attribute browser” is really nice to find what node links to what other node by clicking on a link (so you don’t have to pan and check, etc)
- Export to PDF often works best as an output display (though SVG is also supported)
- If you select an edge and then Ctrl-left-click on the edge, an edge “handle” will appear. This handle can be used to change the shape of the line
- Use the CSV file to make quick modifications, and then check it with the Organic layout
- A convenient way to check the propagation of a network is to select a node, then click on Ctrl+6 again and again (Ctrl+6 selects neighborhood nodes of a selected node, so it “shows” you the network created by a node and its relationships)
- If you want to analyze a sub-graph, search for a node, then press a couple of times on Ctrl+6, then create another graph from that selected node (File → New → Network → from selected node)
- If you begin to see slow performance, then save and re-load your session; there appears to be some memory leaks in the program
- Also, for very large graphs, avoid repeated use of certain layouts (Hierarchy, Orthogonal, etc.) that take very long times to re-draw.
- Degree and Edge count are the same metric.

APPENDIX H: UMBEL WEB SERVICES

These first set of UMBEL Web services provide online demo sandboxes, and descriptions of what they are about and their API documentation. The first 11 services are:

- [Find Subject Concepts](#) — look up via name, URI or alias one of the 21,000 UMBEL subject concepts
- [Subject Concept Report](#) — basic information about the subject concept; description, canonical name, and the like
- [Subject Concept Detailed Report](#) — very cool! see below
- [List Sub-Concepts & Sub-Classes](#) — for developers and ontology geeks
- [List Super-Concepts & Super-Classes](#) — for developers and ontology geeks
- [List Equivalent External Classes](#) — for developers and ontology geeks
- [Verify Sub-Class Relationship](#) — for developers and ontology geeks
- [Verify Super-Class Relationship](#) — for developers and ontology geeks
- [Verify Equivalent Class Relationship](#) — for developers and ontology geeks
- [Subject Concepts Explorer](#) — a very nifty, interactive, graph relationships explorer, and
- [Yago Ontology](#) — direct access to the related YAGO system.