a modular software stack and reusable set of components for semantic content management
Semantic Content Management with Apache Stanbol
Semantic Content Management with Apache Stanbol

- **Enhancer**: Extracts Knowledge from parsed Content
- **Entityhub**: Manage Entities and Topics of Interest to your Domain
- **Contenthub**: Semantic Indexing / Search over your - semantic enhanced - Content
- **CMS Adapter**: Sync. your CMS with Apache Stanbol (JCR/CMIS)
- **Ontology Manager**: Manage your formal Domain Knowledge
- **Reasoners & Rules**: Apply Domain Knowledge to improve / validate extracted Information. Refactor / refine knowledge to align it to public schemas such as schema.org
Stanbol Enhancer

```
curl -X POST -H "Accept: text/turtle" -H "Content-type: text/plain" \
--data "The Stanbol enhancer can detect famous cities such as \
    Paris and people such as Bob Marley." \ 
http://localhost:8080/enhancer
```

Enhancement Chain: **default** all 5 engines available
- tika (optional, TikaEngine)
- langid (required, LangIdEnhancementEngine)
- ner (required, NamedEntityExtractionEnhancementEngine)
- dbpediaLinking (required, NamedEntityTaggingEngine)

Extracted entities

<table>
<thead>
<tr>
<th>People</th>
<th>Places</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob Marley</td>
<td>Paris</td>
</tr>
</tbody>
</table>
Enhancement Engines 1/2

- Apache Tika Engine / Metaxa Engine
  - Plain Text extraction; Metadata Extraction; Content Type detection
- Language Detection
- Topic Classification
  - Trainingset / Classifier for your Topics
  - supports hierarchical Classification Schemes
- Named Entity Recognition
  - extracts Persons / Organizations / Places
Enhancement Engines 2/2

- Named Entity Linking
  - Links recognized Entities with Controlled Vocabularies
- Keyword Extraction
  - Label based extraction of Entities
- Refactor Engine
  - Rule based post-processing of Enhancements results
- Integrated “external” Services:

  Zemanta™  GeoNames

apache stanbol™  CALAIS
Domain Specific Enhancement

If you have any of these other conditions, you may need a dose adjustment or special tests to safely take aspirin:

- asthma or seasonal allergies;
- stomach ulcers;
- liver disease;
- kidney disease;

Enhancement Chain: ehealth all 4 engines available

- tika (optional, TikaEngine)
- langid (required, LangIdEnhancementEngine)
- ehealthExtraction (required, KeywordLinkingEngine)
- drugIdExtraction (required, KeywordLinkingEngine)

Extracted entities

<table>
<thead>
<tr>
<th>Diseases</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td>Aspirin</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td></td>
</tr>
<tr>
<td>Polycystic liver disease</td>
<td></td>
</tr>
</tbody>
</table>
Enhancement Chains

- Define how Content is processed by the Enhancer
  - `/enhancer` calls the default Chain
  - use multiple Chains `/enhancer/chain/{name}`
  - call single EnhancementEngines `/enhancer/engine/{name}`

- Some Examples:
Using Stanbol in Web Applications

**HALLO — ANNOTATING CONTENT WITH LINKED DATA**

Before taking Lotrel

You should not use Lotrel if:

- you are allergic to **amlodipine** (Norvasc) or **benazepril** (Lotensin);
- you have ever had **angioedema** (hives or severe swelling of deep skin tissues sometimes caused by allergies);
- you are allergic to any other ACE inhibitor, such as captopril (Capoten), fosinopril (Monopril), enalapril (Vasotec Zestil), moexipril (Univasc), perindopril (Aceon), quinapril (Accupril), ramipril (Altace), or trandolapril (Mavik).

To make sure you can safely take Lotrel, tell your doctor if you have any of these other conditions:

- kidney disease (or if you are on dialysis);
- liver disease;
- heart disease or **congestive heart failure**;
- diabetes; or
- if you are on a low-sodium diet.

FDA pregnancy category C

Lotrel can cause hyperkalemia (high blood potassium levels) and may not be the best choice during pregnancy. Lotrel can cause harm a nursing baby. Do not breastfeed while using your second or third trimesters.

**Demo:** [http://hallojs.org/annotate.html](http://hallojs.org/annotate.html)

**http://viejs.org**

**http://hallojs.org**
We are looking for

- RDFa / Microdata support
  - Knowledge extraction while keeping positioning within the Content
- Entity Disambiguation
  - Entity-Linking + Disambiguation (e.g. by using Solr MLT)
  - Disambiguation of already linked Entities
- More Domain specific Customizations
  - Share as “/demo” with the Stanbol Community!
- <Your> Service as EnhancementEngine
Stanbol Entityhub

- Manage your Entities
  - e.g. Employees, Projects, User Tags, Company Thesaurus

- Used for Entity-Linking by the Enhancer
  - full local Caches

- cache retrieved Entities

- Support for Linked Data

- Support specific Sites via Extensions

Web Resources
- revyu.com
- linkedct.org

Specific Resources
- Clinical Trials

Local Cache
- Entityhub
- dbpedia
- revyu cache
- linkedct cache
- reuters cache

Remote Sites
- Reuters
- News about Entities

Referenced Sites
Stanbol Entityhub

- Manage multiple Entity Source - Referenced Sites
  - Supports fast local Caches using [Solr](http://solr.apache.org) or [Clerezza](http://clerezza.org)
- Query for Entities
  - used by the Stanbol Enhancer
- LDpath [1] support for:
  - graph path retrieval
  - schema translation
  - simple reasoning

```bash
curl -X POST -d "name=lyon&limit=10" \
http://localhost:8080/entityhub/site/dbpedia/find
```

```python
friend-names = foaf:knows/foaf:name

schema:name = rdfs:label[@en];
schema:description = rdfs:comment[@en];
schema:image = foaf:depiction;
schema:url = foaf:homepage;

skos:broaderTransitive = (skos:broader)+;
skos:related = (skos:related | ^skos:related);
```

You can help by

* Integrate with Data Reconciliation Tools
  * Google Refine:
  * Silk: Entity Link discovery Framework

* Support for <your> Dataset
  * direct access via EntityDereferencer implementation
  * provide as Entityhub ReferencedSite (or RDF dump)
Stanbol Contenthub

curl -i -X POST -H "Content-Type:text/plain" \ 
--data "Add your content here" \ 
http://localhost:8080/contenthub/contenthub/store

Enhancer

Configure your Semantic Index Layout

Simple Faceted Search

RESTful API

Semantic Search

Semantic Indexing

Semantic Index

enhanced Content

plain Content

CMS Adapter
Stanbol Contenthub

✦ Add Semantic Search to your CMS
  ✦ RESTful Faceted Search Interface
  ✦ Related Keyword Search using Entityhub, Ontonet or Wordnet

✦ Improve Search by Semantic Indexing
  ✦ Keep using as you Search Engine
  ✦ Use the Stanbol Contenthub for semantic indexing
  ✦ Configure Semantic Indexes by using LDpath
currently in Development

- Enhancer
  - Provide annotated Content
  - plain Content → enhanced Content

- Store
  - Stanbol will keep Indexes in Sync

- Semantic Indexer
  - Support for different Semantic Indexes

- CMS Adapter
  - Your CMS
  - File System

- Apache Solr
  - Apache Jackrabbit
  - clerezza
  - jena
  - CouchDB

coming with Stanbol 0.1.0
follow STANBOL-471
Stanbol Ontology
Manager, Reasoning and Rules

- Manage your Ontologies
  - and use/combine them in Scopes

- Reasoning
  - on volatile Data loaded into a Sessions
  - consistency check / classification / enrichment
  - RDFS, OWL and OWL - 2

- Support for background Jobs
  - for long running reasoning tasks
Stanbol Ontology
Manager, Reasoning and Rules

Ontologies
- SIDER
- DrugBank
- Dailymed
- Life Science

Scopes
- Media
- Creative Commons
- Music Ontology
- LinkedMovie DB schema
- Social Network

Reasoner
Rules

Sessions

Enhancements
Tim's Profile

revyu.com
Stanbol Ontology
Manager, Reasoning and Rules

- Stanbol Rules
  - Recipes: Manage a set of Rules that are executed together
  - Rules are converted to SWRL, Jena Rules or SPARQL CONSTRUCT depending on the available RuleEngine

- Typical Use Cases
  - integrity checks for imported Data
  - harmonize Vocabularies e.g. simple SEO by using schema.org

- Refactor Enhancement Engine
  - allows to execute Recipes on extracted Metadata
Contributions
Welcome

- Share alignment rules across multiple domains
  - Especially with schema.org.
- Benchmarking:
  - How large are the scopes you are managing?
  - Sessions you use in your applications
- Wrap <your> Reasoner/Rule Engine as a Stanbol service
Stanbol Design and Integration Patterns

- Stanbol Components provide
  - RESTful API
  - Java API and OSGI services
- Stanbol Components do NOT depend on each other
  - however they can be easily combined to

Don’t buy everything. Take the Components you Need!
Stanbol Facts

- Mailing List: stambol-dev@incubator.apache.org
- Release: in progress (currently: 0.9.0-incubation RC6)
- Incubation to Apache November 2010
  - based on code developed by the [IKS](http://www.iks-project.eu) project [1]

[1] [http://www.iks-project.eu](http://www.iks-project.eu)