Integrated development environment for business solutions based on knowledge networks for smart data

Semantics 2019 Conference
September 10, 2019

Christian Piepenbrock, Johannes Munk

Klarso GmbH Berlin – www.klarso.com
Where we are headed
Smart Data

How we innovate
Semantic Networks

Why it’s efficient
Integrated Development Environment

What we deliver
Application Cases

What’s next
Outlook
Shift in the Information Landscape

Smart Data

#1 catalyst for business growth

real world relations today

index cards 1900s

tabular data 1960s

big data 2010s
Smart Data for Product Information and Content Delivery

Customers Expect Smart New Experiences

- Product Information
  - Relevant offerings
  - Personalized service
  - Comprehensive info
  - Requirements are met
  - Product configuration

- Content Delivery
  - Highest quality
  - Content interaction
  - Social context
  - Deep search
  - Information advantage

Semantzics 2019
10.09.2019
Smart Data to Create Your Top-Down Artificial Intelligence

- Realized in *klar:suite*
  - Semantic network knowledge base
- Establish a top-down AI
  - Fully explainable high level reasoning

Automatic cloning, versioning, validity checks through structural understanding

Automatic reasoning of implications, detection of contradictions or structural analogies

Automatic explanation of data relationships with connection to natural language

10.09.2019
Semantzics 2019
Where we are headed
Smart Data

How we innovate
Semantic Networks

Why it’s efficient
Integrated Development Environment

What we deliver
Application Cases

What’s next
Outlook
Semantic Network Problems and Claims

**Sustainability**
- traditionally: fixed design (ontology)
- usually designed for specific applications and not content
- neglects what you might need in the future
- you need a **natural representation** of information

**Quality & Flexibility**
- generic tools cumbersome to use
- missing support for specific tasks
- errors through manual curation unavoidable
- effort scales badly with complexity
- you need **custom and agile tools**

**Scalability**
- Useful semantic networks get big really fast
- it is hard to keep the overview
- you need **smart data** that can take care of itself
- we've got 6 revolutionary improvements...

10.09.2019 Semantzics 2019
Innovation: We Revolutionize Semantic Networks in 6 Ways

1. Representation power
   Continuous abstraction, context awareness, new relation types

2. Flexibility and agility
   Scripted user interface and other modules with continuous live deployment

3. User interaction
   Specialized tools to visualize, edit, curate data; workbench with workflows

4. Scalability and speed
   Data in proprietary decentralized fast memory-mapped database

5. Linking external data
   Master data management to map/link/transform/enrich/sync external data

6. System integration
   Semantic network embedded into a complete system of integrated modules
Innovation: We Revolutionize Semantic Networks in 6 Ways

1. Representation power

- relations between relations
- more than binary relations
- integrated ontology: relations between types and entries

2. Smart for: self-maintenance and sustainability, logic, references, context definition

3. Representation power

4. Representation power

5. Representation power

6. Representation power
Innovation: We Revolutionize Semantic Networks in 6 Ways

2 Flexibility and Agility

Semantic Network

- powerful representation allows natural structure
- organic evolution of semantic relations through successive enrichment
- sustainable through shared or generated lookup actions to query related entities – avoiding code redundancy.

Smart for: agile adaption of data design to changing requirements.
Innovation: We Revolutionize Semantic Networks in 6 Ways

1. User interaction
2. Fully customizable set, position and size of widgets/views
3. Setup can be stored and loaded to support multiple workflows
4. On selection of an entry every widget populates itself accordingly
5. Different aspects seen in parallel, separation of concerns
6. Outline, text 1, text 2, history, search

Type 1 details
Type 2 details
Type 3 details
Import Report
Search results
Search filters

10.09.2019
Innovation: We Revolutionize Semantic Networks in 6 Ways

4. Scalability and speed

- Data stored in **proprietary, decentralized, fast memory-mapped database**
- Makes interactive views with **100k's** of requests possible
- **Efficient** manipulation of big data sets

**Smart** for: rich sustainable datasets, distributed collaboration
Where we are headed
Smart Data

How we innovate
Semantic Networks

Why it’s efficient
Integrated Development Environment

What we deliver
Application Cases

What’s next
Outlook
**klar:suite – software architecture**

**Goal:** Affordable and sustainable custom tailored tooling for every project.

- **Enduser:** Efficient workflows. Ability to react quickly to changed requirements.
- **Developer:** Fast coding. Less pressure to generalize. Happy customers.

**Powerful framework on semantic net**

- Scripting based on **JavaScript**
- **High performance** C++ core
- **Cross platform** deployment with Qt
- **Model-Views** & HTML5-WYSIWYG with direct read-write access
- Central **event hub** on entries to interconnect widgets and maintain separation of concerns
- **Notifications** for local and remote actions
- **Reusability** through central solution repository and widget inheritance

**Scripted studio setup**

- **Scripted user interface** and functions
- Scripts are part of semantic network => **seamless live deployment**
**klar:suite – framework modules**

**klar:base – semantic database**
entries, types, type inheritance, attributes, relations, n-adic, relations between relations, lists, sets, multi-lingual; *smart*: versioning, ownership.

**klar:widgets**: enhanced widgets for table and treeviews, full range of Qt5 widgets.

**klar:grammar – parsing**
Link-Grammar and spaCy for semantic data extraction from natural language.

**klar:search**: Concept based custom ranking, autocomplete, phonetic reduction, fulltext index.

**klar:server/API**: daemon with full access to semantic network. Executes same JavaScripts as Studio.

**klar:file**: XML; PDF output through HTML5; Excel-xlsx; Word-docx; csv/tsv; Arbitrary (binary) formats.

**klar:sync**: synchronization of user instructions. Offline capable.
How we Deliver
In-House and Cloud Services

Integrate
Import or sync other data sources like ERP, XML content, CRM.

Semantic Network
Distributed semantic network
Instructions synced through SQL server
Export (e.g. XML, PDF, print prepress)
Live or offline on PC or smartphone.

Cloud Services
Connect web server and services by API.

Web & App Access
Deliver HTML or API to clients.
Tooling Perspective

**klar:suite**
- standard modules in agility centered framework
- specialized custom built user interfaces
- ... enhanced over time

**standard software**
- e.g. Excel for tables, iViews for graphs
- no deeper "understanding" of data structure
- no custom editors and views
- **risks**: data inconsistencies, high manual effort

**fully custom built software**
- high development costs
- often dead end in terms of architecture
- know how management critical
- **risks**: data silos, lack of agility
A *klar:suite* Project
Flexible and Agile

**Insights & ideate**

**Prototype (optional)**
Create a working prototype with real data set. Validate utility, user interactions, services.

**Define & develop MVP**
Build data representation & semantic network. Create custom editors, input, output, search.

**Release early**
Roll out earliest MVP. Enable customer to engage with data, search, output generation.

**Co-create expansions**
Build new features based on need & experience. Co-create detailed data interaction and delivery.

**Remain flexible**
Future proof: at any time adapt user interface & data representation through simple scripts.
Where we are headed
Smart Data

How we innovate
Semantic Networks

Why it’s efficient
Integrated Development Environment

What we deliver
Application Cases

What’s next
Outlook
Smart Data Customers

- **Stiftung Warentest**
  - Consumer organization
  - Product information management for "Medicines in Test"

- **Forklift manufacturer**
  - Product information management for European markets

- **DE GRUYTER**
  - Scholarly publisher
  - Medical reference platform "Pschyrembel"

- **Online news publisher**
  - Automatic language analysis for news articles

- **Kaiser + Kraft**
  - B2B supplies
  - Vendor product data onboarding
Product information management for medical drugs (Stiftung Warentest, Project “Medicines in Test”)

- The project
  - End consumer web platform with details and test results on the 8,000 most used drugs by consumer organisation “test”
- The klar:suite system
  - Semantic database for medical drugs and all general diseases semantically interlinked for disease, symptoms, treatments, mechanisms of action, molecule, delivery form, trade names, packages, side effects, etc.
- The features
  - Fully featured editorial backend system to edit and semantically enrich the data
  - Regular automatic update through import of externally provided drug database
  - XML/HTML export to frontend web platform
Product information management for medical drugs (Stiftung Warentest, Project “Medicines in Test”)

- Flexible data model allows the agile development of visionary new prototypes
- Direct interaction with the semantic network through custom-tailored dynamic data views
- Highest quality content through automatic data consistency checks and reports
- Fine granular data representation of entities and relations to avoid redundant editing
- Lightweight reports allow prompt responses to marketing enquiries
- Smart heuristics for external data integration and self-maintenance of expiring datasets
Product information management for medical drugs
Versions, revisions, approval

- **Diff view**
  - Doc diff view between any two texts
  - Individual approval/rejection of changes

- **Validation**
  - Automatic rule-based validation of docs

- **Approval**
  - Custom approval workflows as needed, e.g. for docs, modules, translations, etc.

- **Automatic revisions**
  - Automatic processes for updates/revisions, e.g. price updates, aggregate changes, etc.

- **On-demand generation**
  - Automatic document generation on demand, e.g. template-based live views
Status
- Custom status for docs, modules, transl., etc.
- Tracking of changes, timestamps, editors

Workflow & rights
- Custom workflows, pre-defined or click&go
- Differentiated user views and rights
- Preferred: allow changes but require approval

Versions
- Full version history, diff views, cloning of docs
- Branching (e.g. „in production fix“ vs. „beta“)

Collaboration
- Fully featured editor seat: *klar:studio*
- External search, review, input: web view
- External editors through Word/Excel im-/export
Product information management automotive (Crown Lift Trucks, forklift manufacturer)

- The project
  - PIM for all model series and options, with prices for all European markets and channels
- The klar:suite system
  - Integrate highly complex product data for all model variants, options, and configuration rules from ERP and sales systems
  - Deliver PDF price lists with all product rules
- The features
  - Full control over product data, configuration options, and prices in-house with klar:studio
  - Clean single point of truth for all product data
  - Automated generation of specific price lists for all markets, currencies, and distr. channels
  - Richest possible dataset for next generation marketing and sales solutions
Organization of text modules
- Custom doc types and structures, e.g. service, installation, maintenance, training, mktg.

Assembly of modules to documents
- Dynamic generation from semantic analysis
- Insert text modules with drag & drop
- Flexible module granularity: word - paragraph

Document templates
- Start from templates or cloned document
- Rules enforceable; variants, e.g. print vs. web

Automatic validation: content & consistency

Permanent backlink to source data element
- Exact links for modules, terms, numbers, etc.
- Edits in preview modify the source element
Product information management automotive
Master data management

- Semi-interactive data mapping
  - Map different data sources
  - Automatic mapping with granular heuristics
  - Visualize mapping in table or layout view for review and revision
  - Inverted view: see multiple or missing mappings

- Mapping reports
  - Specialized reports for mapping errors: mismatches, missing data, ...

- Master data management
  - Single point of truth for each data point in klar:suite or externally
  - Store foreign keys as relations to imported / external data
Medical reference platform for physicians (De Gruyter, scholarly publisher, „Pschyrembel“)

- The project
  - German medical reference dictionary, “online first” for web, app, and print

- The klar:suite system
  - Digital delivery platform for all channels
  - CMS system for smart authoring and semantic enrichment
  - Flexible data model with inter-linked enriched content (articles, references, glossary, etc.)

- The features
  - XML structure completely re-factored into a system that exceeds XML capabilities
  - Multiple publications integrated
  - Workflows to support editors with more than 250 external authors
Medical reference platform for physicians

Semantic search

- Search
  - Systematic access by specialty or ICD-10 code
  - Deep search for concepts
  - Customized search
  - Autocomplete, result snippets
  - Best hits with abstract
  - Search across multiple publications
  - Faceted search results with tree view
  - Search term highlighting
  - Full text search
  - Concept links
Medical reference platform for physicians

Article display

- Article display
- Navigation by section headings
- Concept panel: lex. Info, synonyms, ethym., translations, ICD-10 codes
- Terminology or links: preview on mouse-over
- Media assets: images, video, tables w. caption
- Breadcrumbs, abstract
- Search within article
- Paywall
- Additional premium content
Automatic language analysis for news articles (Golem media, online news company)

- The project
  - Artificial intelligence analysis of incoming press releases for editors, archive, and topic aggregation with concept tagging

- The klar:suite system
  - Press release processing system that analyzes texts by natural language analysis
  - Discover and extract topics, entities (like people, products), relations (key statements)

- The features
  - The flexible data model gets continuously adapted through learning from texts, through training new grammatical structures, by externally validating extracted information
  - Golem now has an automatic press release database, tags it semantically, validates it using editors and crowd sourcing
Automatic language analysis for news articles

Natural language extraction

- Automatic entity extraction from natural language
- Integration into automatic workflows through APIs
- Validation of concepts and relations through editorial review and crowd-sourcing
Where we are headed
Smart Data

How we innovate
Semantic Networks

Why it’s efficient
Integrated Development Environment

What we deliver
Application Cases

What’s next
Outlook
Other applications and outlook

- Industry
  - offer generation
  - machine and product configuration
  - technical documentation
  - quality and regulatory

- Legal tech / insurance
  - complex document analysis
  - natural language processing
  - contract structuring and drafting

- Terminology management
  - term relations
  - domain knowledge management

- Automatic software code generation
  - based on natural language of developer comments

Make software development agile,
Make workflows flexible,
Make data smart
Thank you

Kontakt:
Christian Piepenbrock, Geschäftsführer
Klarso GmbH
Schwartzkopffstr. 7 a, 10115 Berlin
Email piepenbrock@klarso.com
Phone +49-173-2063907