|  | **Knowledge Graph Extractor (Hella)** |
| --- | --- |
|  |  |
| **Online team meeting** | https://fau.zoom-x.de/j/67111681334?pwd=LzdBM3lXeXhPTEtWL3lGUnFqbTAzZz09 |
|  |  |
| **Production system (if any)** | ... |
| **Test system (if any)** | ... |
|  |  |
| **GitHub repository** | https://github.com/amosproj/amos2024ss05-knowledge-graph-extractor |
| **GitHub feature board** | https://github.com/orgs/amosproj/projects/56/views/2 |
| **GitHub impediments backlog** | ... |
|  |  |
| **Team T-shirt (white)** | ... |
| **Team T-shirt (black)** | https://www.shirtinator.de/s/pYjJO4qcR3u9lSKbgQdyiw |
|  |  |
| **Additional materials** | ... |
|  |  |
| **Team maling list** | oss-amos-proj5@lists.fau.de |
|  |  |
| **Project notes** | https://docs.google.com/document/d/1OlcUP4HQkkPC0CHgqSkg_3PN46-r5dKddH-FLOZh-uo/edit |
| **Questions for Industry Partner** | https://docs.google.com/document/d/1DsWOP9P-WQFEtSQg0-_vFGQ1AePdJhCbIhrcKZ2X8oE/edit |

| Last Name | First Name | GitHub User Name | Email Address |
|---|---|---|---|
| Kuo | Irene | kuoirene | kuo.irene.y@gmail.com |
| Greiner | Rebecca | RebeccaGreiner | rebecca.greiner@fau.de |
| Rauscher | Nikolas | nikolas-rauscher | nikolas.rauscher@gmail.com |
| Ozseker | Irem | iremozs | iremozseker@gmail.com |
| Müller | Hanna | hanna-212 | hanna.mueller@fau.de |
| Fabian Borges | Filipe Alexandre | borges-filipe | filipe.af.borges@gmail.com |
| Kotini | Kristi | kristikotini | kristi.kotini@fau.de |
| Bhesaniya | Yash | yashbhesaniya | yashbhesaniya1999@gmail.com |
| Ramesh | Sandeepkumar | Sandeep-kumar-Ramesh | sandeepkumar.ramesh@fau.de |
| Hoffmann | Florian | get4flo | f.hoffmann@campus.tu-berlin.de |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| # | Meeting Day | Product Owners | Software Developer | Release Manager | Scrum Master | Comment |
|---|---|---|---|---|---|---|
| 1 | 2024-04-17 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | |
| 2 | 2024-04-24 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | |
| 3 | 2024-05-01 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | |
| 4 | 2024-05-08 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | |
| 5 | 2024-05-15 | Irene Kuo, Rebecca Greiner | Everyone else | Kristi Kotini | Hanna Müller | |
| 6 | 2024-05-22 | Irene Kuo, Rebecca Greiner | Everyone else | Nikolas Rauscher | Hanna Müller | |
| 7 | 2024-05-29 | Irene Kuo, Rebecca Greiner | Everyone else | Sandeepkumar Ramesh | Hanna Müller | Mid-term due |
| 8 | 2024-06-05 | Irene Kuo, Rebecca Greiner | Everyone else | Yash Bhesaniya | Hanna Müller | |
| 9 | 2024-06-12 | Irene Kuo, Rebecca Greiner | Everyone else | Florian Hoffmann | Hanna Müller | |
| 10 | 2024-06-19 | Irene Kuo, Rebecca Greiner | Everyone else | Filipe Borges | Hanna Müller | |
| 11 | 2024-06-26 | Irene Kuo, Rebecca Greiner | Everyone else | Kristi Kotini | Hanna Müller | |
| 12 | 2024-07-03 | Irene Kuo, Rebecca Greiner | Everyone else | Yash Bhesaniya | Hanna Müller | |
| 13 | 2024-07-10 | Irene Kuo, Rebecca Greiner | Everyone else | Irem Ozseker | Hanna Müller | |
| 14 | 2024-07-17 | Irene Kuo, Rebecca Greiner | Everyone else | Sandeepkumar Ramesh | Hanna Müller | Demo day! |
| 15 | 2024-07-24 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | Retrospective |

| | |
|---|---|
| **Goals** | 1. Finish tasks for each sprint on time. |
| | |
| **Meeting norms** | 1. Be on time! (send a msg in WhatsApp if you'll be late)<br>2. Show up (unless deathly sick)<br>3. Try to participate actively |
| | |
| **Working norms** | 1. Good comments/documentation of work so everyone can follow easily.<br>2. Don't do everything the day before it's due.<br>3. Reach out if you have questions, help each other out! |
| | |
| **Coordination norms** | 1. Make it clear on the feature board what you're working on.<br>2. If you're overwhelmed, communicate so we can reassign tasks. |
| | |
| **Communication norms** | 1. Create WhatsApp group and reach out for questions and concerns there first (informal quick chats)<br>2. Discord for screenshots, code concerns, one point of reference for project items. |
| | |
| **Consideration norms** | 1. Be kind to each other. |
| | |
| **Cont. improvement norms** | 1. Have a retrospective after each sprint. |
| | |
| **Rewards** | Everyone bring your own treat and we can have a celebratory meeting at the end! |
| | |
| **Sanctions** | If you're more than 5min late without notice, 1pushup per minute late is owed. |
| | |
| **Signatures** | |
| | |
| Scrum Master | Hanna Müller |
| Product owner | Irene Kuo |
| Product owner | Rebecca Greiner |
| Software developer | Nikolas Rauscher |
| Software developer | Irem Ozseker |
| Software developer | Yash Bhesaniya |
| Software developer | Filipe Borges |
| Software developer | Kristi Kotini |
| Software developer | Florian Hoffmann |
| Software developer | Sandeepkumar Ramesh |

| Product Vision | Project Mission |
|---|---|
| An AI-powered chatbot that helps any user query and extract knowledge from uploaded document(s). Through generating knowledge graphs from a corpus of text, information and knowledge is organized in a smarter way that is able to reveal different insights that may not have been noticed before.<br>The knowledge graph will include communities of concepts and can be used to uncover insights and links between seemingly disconnected concepts. Through querying knowledge graphs, users can more quickly gather the correct information and potentially gain additional understandings that are not noticeable without the graph communities. | The mission of this project is to create a MVP for the knowledge graph generation in order to visually see clusters of information and how they're linked. The knowledge graph will include a basic search function to query information.<br><br>Core functionality will be ingesting user document(s), processing the data and extracting relationship entities through the use of LLMs, building and storing the knowledge graph, an interactive visual representation of the knowledge graph, and a basic search function for entities in the knowledge graph. |

| Term | Definition |
|---|---|
| Knowledge Graph (KG) | A knowledge base that uses a graph structure to represent the data with nodes as objects and edges as relationships between the nodes. |
| Technical document | A piece of written content that provides detailed information, instructions, or explanations about a specific technical subject, product, or process. |
| ASPICE | The latest standard for best practices in automotive software. |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

| Sprint # | Sprint goal |
|---|---|
| 1 | None |
| 2 | None |
| 3 | None |
| 4 | Optional |
| 5 | Finish all basic components/functions in preparation for connecting them all for the end-to-end functionality (upload -> knowledge graph visualization). |
| 6 | MVP for mid-project with upload through 1st basic visualization of knowledge graph |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| | |
| | |
| | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| **Release** | | | | | | |
| | | | | | | |
| **Total** | | | 124 | 124 | | |
| | | | | | | |
| **Sprints** | | | | | | |
| | | | | | | |
| 1 | **Getting started** | | 0 | 124 | 0 | 124 |
| 2 | **Define technologies, create software architecture and user interface design** | | 13 | 124 | 13 | 124 |
| | | | | | | |
| 3 | **Setup project environment** | | 16 | 111 | 14 | 111 |
| 4 | **Ingestion of documents and LLM setup with POC of data processings** | | 32 | 95 | 31 | 97 |
| 5 | **Preparation of individual functions to connect for the MVP** | | 32 | 63 | 32 | 66 |
| 6 | **MVP for mid-project with upload through 1st basic visualization of knowledge graph** | | 31 | 31 | 31 | 34 |
| | **Sum** | | | | | |
| | | | | | | |
| **Features** | | | | | | |
| | | | | | | |
| 1 | **Getting started** | Setup feature board | n/a | | n/a | |
| | | | | | | |
| 2 | **Define technologies, create software architecture and user interfac** | Team logo | n/a | | n/a | |
| | | Create software architecture overview | 5 | | 5 | |
| | | Design user interface | 8 | | 8 | |
| | | | | | | |
| 3 | **Setup project environment** | Set up initial project environment (backend excluding LLM container) | 8 | | 8 | |
| | | | | | | |
| 4 | **Ingestion of documents and LLM setup with POC of data processings** | PDF parsing into text | 5 | | 3 | |
| | | Text to .json chunks | 3 | | 3 | |
| | | Interface setup | 3 | | 3 | |
| | | Allow user to upload PDF document(s) | 5 | | 5 | |
| | | POC: Graph visualization | 5 | | 5 | |
| | | Setup Mistral locally: documentation | 3 | | 2 | |
| | | POC: Prompt template for LLM | 3 | | 5 | |
| | | Syntax checking for JSON and converting to graph format | 5 | | 5 | |
| | | | | | | |
| 5 | **Preparation of individual functions to connect for the MVP** | Update software architecture diagram and documentation | 1 | | 1 | |
| | | Prepare LLM setup for dev team | 2 | | 2 | |
| | | POC: combine graph pieces with LLM | 8 | | 8 | |
| | | Create record in database | 5 | | 5 | |
| | | LLM function | 3 | | 3 | |
| | | Generate graph button | 5 | | 5 | |
| | | CORS implementation | 3 | | 3 | |
| | | CI/CD improvements | 2 | | 3 | |
| | | HW: Build process video | 3 | | 2 | |
| | | | | | | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| | | | | | | |
| 6 | **MVP for mid-project with upload through 1st basic visualization of knowledge graph** | Save graph data | 8 | | 8 | |
| | | User documentation | 3 | | 3 | |
| | | Connecting components from documents to entities | 5 | | 5 | |
| | | Deploy documentation | 3 | | 3 | |
| | | Connect entities to graph data | 5 | | 5 | |
| | | Generate graph visualization from database | 5 | | 5 | |
| | | Technical/design documentation | 2 | | 2 | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| **Release** | | | | | | |
| | | | | | | |
| **Total** | | | 181 | 181 | | |
| | | | | | | |
| **Sprints** | | | | | | |
| | | | | | | |
| 1 | **Getting started** | | 0 | 181 | 0 | 181 |
| 2 | **Define technologies, create software architecture and user interface design** | | 13 | 181 | 13 | 181 |
| 3 | **Setup project environment** | | 16 | 168 | 14 | 168 |
| 4 | **Ingestion of documents and LLM setup with POC of data processings** | | 32 | 152 | 31 | 154 |
| 5 | **Preparation of individual functions to connect for the MVP** | | 32 | 120 | 32 | 123 |
| 6 | **MVP for mid-project with upload through 1st basic visualization of knowledge graph** | | 31 | 88 | 31 | 91 |
| 7 | **Streamline UX and work on additional KG generation tasks** | | 26 | 57 | 26 | 60 |
| 8 | **Update UI and KG fine-tuning** | | 31 | 31 | | |
| 9 | **Enhance graph visualization and LLM-usage** | | | | | |
| 10 | **Additional graph search and KG fine-tuning** | | | | | |
| 11 | **Bug fixes and last fine-tuning tasks** | | | | | |
| 12 | **Finish final project release and prepare for demo day** | | | | | |
| | **Sum** | | | | | |
| | | | | | | |
| **Features** | | | | | | |
| | | | | | | |
| 1 | **Getting started** | Setup feature board | n/a | | n/a | |
| | | | | | | |
| 2 | **Define technologies, create software architecture and user interface design** | Team logo | n/a | | n/a | |
| | | Create software architecture overview | 5 | | 5 | |
| | | Design user interface | 8 | | 8 | |
| | | | | | | |
| 3 | **Setup project environment** | Set up initial project environment (backend excluding LLM container) | 8 | | 8 | |
| | | | | | | |
| 4 | **Ingestion of documents and LLM setup with POC of data processings** | PDF parsing into text | 5 | | 3 | |
| | | Text to .json chunks | 3 | | 3 | |
| | | Interface setup | 3 | | 3 | |
| | | Allow user to upload PDF document(s) | 5 | | 5 | |
| | | POC: Graph visualization | 5 | | 5 | |
| | | Setup Mistral locally: documentation | 3 | | 2 | |
| | | POC: Prompt template for LLM | 3 | | 5 | |
| | | Syntax checking for JSON and converting to graph format | 5 | | 5 | |
| | | | | | | |
| 5 | **Preparation of individual functions to connect for the MVP** | Update software architecture diagram and documentation | 1 | | 1 | |
| | | Prepare LLM setup for dev team | 2 | | 2 | |
| | | POC: combine graph pieces with LLM | 8 | | 8 | |
| | | Create record in database | 5 | | 5 | |
| | | LLM function | 3 | | 3 | |
| | | Generate graph button | 5 | | 5 | |
| | | CORS implementation | 3 | | 3 | |
| | | CI/CD improvements | 2 | | 3 | |
| | | HW: Build process video | 3 | | 2 | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| 6 | **MVP for mid-project with upload through 1st basic visualization of knowledge graph** | Save graph data | 8 | | 8 | |
| | | User documentation | 3 | | 3 | |
| | | Connecting components from documents to entities | 5 | | 5 | |
| | | Deploy documentation | 3 | | 3 | |
| | | Connect entities to graph data | 5 | | 5 | |
| | | Generate graph visualization from database | 5 | | 5 | |
| | | Technical/design documentation | 2 | | 2 | |
| | | | | | | |
| 7 | **Streamline UX and work on additional KG generation tasks** | Linting/Formatting | 3 | | 3 | |
| | | Create landing page | 3 | | 3 | |
| | | View list/table of existing knowledge graphs | 5 | | 5 | |
| | | Create new knowledge graph button (link to current user flow) | 2 | | 2 | |
| | | Delete uploaded document from upload screen | 3 | | 3 | |
| | | Refine .json extraction from LLM results | 3 | | 3 | |
| | | Refine graph connections | 3 | | 3 | |
| | | Ordering size of nodes for graph visualization | 3 | | 3 | |
| | | Remove JanusGraph | 1 | | 1 | |
| | | | | | | |
| 8 | **Update UI and KG fine-tuning** | Update user interface pages to design theme | 3 | | | |
| | | Allow users to delete knowledge graph | 2 | | | |
| | | After LLM results, eliminate duplicate entities | 5 | | | |
| | | POC: Query knowledge graph (to help with evaluating it) | 8 | | | |
| | | Link entities to page | 3 | | | |
| | | Improve visualization based on different node sizes | 3 | | | |
| | | Experiment with different approaches | 5 | | | |
| | | View knowledge graph from table list | 2 | | | |
| | | | | | | |
| 9 | **Enhance graph visualization and LLM-usage** | Graph display text/node color + less overlapping of nodes | | | | |
| | | Find way to improve performance time | | | | |
| | | Finetuning of prompt template and ontology - make it more abstract and more concise | | | | |
| | | Look deeper into centrality measures for making network more concise | | | | |
| | | force-based algorithm for node positions | | | | |
| | | Split view - to show more information on left side, graph on right | | | | |
| | | "Generate" button to link to generate graph for documents that only have been uploaded | | | | |
| | | Refactoring: "delete uploaded document" button | | | | |
| | | Refactoring: clean up code base | | | | |
| | | Clustering of nodes / topic modeling: attributes | | | | |
| | | Progress bar for upload page | | | | |
| | | | | | | |
| 10 | **Additional graph search and KG fine-tuning** | Follow up on POC of graph search - implement functionality for user | | | | |
| | | Clustering of nodes / topic modeling: coloring (#106) | | | | |
| | | Another iteration of knowledge graph prompt fine-tuning (#104) | | | | |
| | | Graph search: return pages and text chunks (#105) | | | | |
| | | If graph structure analysis/centrality measures haven't been applied, choose what's best and apply them (#121) | | | | |
| | | Draggable nodes | | | | |
| | | Bug fixes | | | | |
| | | | | | | |
| | | | | | | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| 11 | Bug fixes and last fine-tuning tasks | Demo day slide | | | | |
| | | Demo day video | | | | |
| | | Bug fixes | | | | |
| | | Final iteration of knowledge graph fine-tuning | | | | |
| | | Graph search: see if there's a way to make returned results more concise | | | | |
| | | Clean-up codebase | | | | |
| | | | | | | |
| 12 | Finish final project release and prepare for demo day | Finalize user, (technical) design, and build/deploy documentation | | | | |
| | | End-to-end testing of application and features | | | | |
| | | Finalize demo day workflow | | | | |
| | | Final project report | | | | |
| | | | | | | |
| | | | | | | |

| # | Feature Definition of Done | Sprint Release Definition of Done | Project Release Definition of Done |
|---|---|---|---|
| | Acceptance criteria is satisfied | Release tag candidate builds and deploys properly | Project builds and deploys properly |
| | Pull request to dev branch | All previous working features should still work properly | Proper documentation on how to use and build the project is done |
| | Code-reviewed by peer | | |
| | Approve code and merge into dev branch | | |
| | Automated tests are run and passed | | |
| | When necessary, update software architecture diagram/documentation and bill of materials | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| Type | Link / reference |
|------|------------------|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

| | Context | Name | Version | License | Comment |
|---|---|---|---|---|---|
| 1 | PDF text extraction | pypdf | v4.2.0 | new BSD | pdf2text |
| 2 | Splitting text into chunks | LangChain | v0.1.17 | MIT | |
| 3 | LLM | Mistral-7B Instruct | v0.2 | Apache 2.0 | No longer in use! |
| 4 | Locally running LLM | Ollama | v0.1.33 | MIT | No longer in use! |
| 5 | Working with the data | pandas | v2.2.2 | new BSD | |
| 6 | Generating graph from data | NetworkX | v3.3 | new BSD | python package, this version requires Python 3.10, 3.11, or 3.12. |
| 7 | Upload documents | Filepond | 4.31.1 | MIT | |
| 8 | Network service | Axios | 1.6.8 | MIT | |
| 9 | Visualization | D3 | v7.9.0 | ISC License (functionally equivalent to the BSD 2-Clause and MIT licenses) | |
| 10 | Visualization | Cytoscape | 3.10.2 | MIT | |
| | Visualization | Vis.js | v9.1.9. | Apache 2.0 / MIT | |
| | Visualization | G6 | 4.8.24 | MIT | |
| | Operational database | Postgres | 16.2 | PostgreSQL license (similar to MIT) | |
| 11 | LLM (more powerful option) | Gemini | 1.5 | Google API Terms of Service | might switch to this LLM from the original one |
| | LLM (more powerful option) | llama3 | llama3-8b-8192 | Groq API Terms of Service | |
| | | | | | |
| | | | | | |
| | | | | | |

| Last Name | First Name | Value | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ramesh | Sandeepkumar | 3 | | **3.00** | **OK** | | | |
| Hoffmann | Florian | 3 | | | | | | |
| Rauscher | Nikolas | 3 | | | | | | |
| Ozseker | Irem | 3 | | | | | | |
| Bhesaniya | Yash | 3 | | 0 | No size | | | |
| Fabian Borges | Filipe Alexandre | 3 | | 1 | Trivial size | | | |
| Kotini | Kristi | 3 | | 2 | Small size | | | |
| | | | | 3 | Medium size | | | |
| | | | | 5 | Large size | | | |
| | | | | 8 | Very large size | | | |
| | | | | 13 | Too large (size) | | | |
| | | | | | | | | |
| **How to play planning poker** | | | | | | | | |
| | | | | | | | | |
| 1. Everyone type their number into their value field, don't hit return yet | | | | | | | | |
| 2. Someone, perhaps a product owner, count down 3.. 2.. 1.. | | | | | | | | |
| 3. Then, everyone hit return to submit their value | | | | | | | | |
| | | | | | | | | |