| Project Name | 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-_vFGQ1AePdJhCblhrcKZ2X8oE/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 | Irem Ozseker | Hanna Müller | |
| 12 | 2024-07-03 | Irene Kuo, Rebecca Greiner | Everyone else | Kristi Kotini | Hanna Müller | |
| 13 | 2024-07-10 | Irene Kuo, Rebecca Greiner | Everyone else | Yash Bhesaniya | Hanna Müller | |
| 14 | 14 | Irene Kuo, Rebecca Greiner | Everyone else | Sandeepkumar Ramesh | Hanna Müller | Demo day! |
| 15 | 1900-01-20 | Irene Kuo, Rebecca Greiner | Everyone else | n/a | Hanna Müller | Retrospective |
| | 1900-01-27 | | | | | |
| | | | | | | |
| | | | | | | |

| | | |
|---|---|---|
| **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 |
| --- | --- |
| WIP | |
| An AI-powered chatbot that helps any user query and extract knowledge from any upload document(s). Through the use of knowledge graphs, different insights that may not have been noticed before can be used to answer queries<br><br>reveal insights into unprecedented interdisciplinary relationships that can be used to answer<br>queries,<br><br>to understand how important a concept (node) is to the body of work. We can analyse connected and disconnected sets of concepts, or calculate communities of concepts for a deep understanding of the subject matter. We can understand links between seemingly disconnected concepts.<br><br>meaningful entities, consistent entities, resilience in parsing, categorization of entities, implied relations<br><br>(Example: The Flowers social network helps flower enthusiasts world-wide to connect with each other and enjoy following their favorite hobby online. Centered on showing and rating favorite flower photos, it inspires growing and presenting ever more | The mission of this project is to create a MVP for the knowledge graph generation. The knowledge graph will be used by the AI-chatbot to query information. 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, and a interactive visual representation of 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. |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

| Sprint # | Sprint goal |
|---|---|
| 1 | None |
| 2 | None |
| 3 | None |
| 4 | Optional |
| 5 | Write your sprint goal here |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| | |
| | |
| | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Release** | | | | | | |
| | | | | | | |
| **Total** | | | 29 | 29 | | |
| | | | | | | |
| **Sprints** | | | | | | |
| | | | | | | |
| 1 | **Getting started** | | 0 | 29 | 0 | 29 |
| 2 | **Define technologies, create software architecture and user interface design** | | 13 | 29 | 13 | 29 |
| 3 | **Setup project environment** | | 16 | 16 | 8 | 16 |
| 4 | **Ingestion of documents (LLM + POC stuff)** | | | | | |
| 5 | | | | | | |
| ... | | | | 0 | | 8 |
| | | | | | | |
| **Features** | | | | | | |
| | | | | | | |
| **1** | **Getting started** | Setup feature board | n/a | | n/a | |
| | | | | | | |
| **2** | **Define technologies, create software architecture and user interfa** | 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 (LLM + POC stuff)** | Text to .json chunks | 5 | | | |
| | | PDF paring into text | 3 | | | |
| | | Allow user to upload PDF document(s) | 5 | | | |
| | | Set up LLM container on Docker | 3 | | | |
| | | ... | | | | |
| | | | | | | |
| | | | | | | |
| 5 | | | | | | |

| Sprint | Goal | Feature Name | Est. Size | Est. Remaining | Real Size | Real Remaining |
|---|---|---|---|---|---|---|
| | | | | | | |
| **Release** | | | | | | |
| | | | | | | |
| **Total** | | | 0 | 0 | | |
| | | | | | | |
| **Sprints** | | | | | | |
| | | | | | | |
| 1 | | | 0 | 0 | 0 | 0 |
| 2 | | | 0 | 0 | 0 | 0 |
| 3 | | | 0 | 0 | 0 | 0 |
| ... | | | | 0 | | 0 |
| | | | | | | |
| **Features** | | | | | | |
| | | | | | | |
| **1** | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| **2** | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| **3** | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

| # | 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 | |
| 4 | Locally running LLM | Ollama | v0.1.33 | MIT | |
| 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 | Graph database | JanusGraph | v1.0 | Apache 2.0 | |
| 8 | Upload documents | Filepond | 4.31.1 | MIT | |
| 9 | Network service | Axios | 1.6.8 | MIT | |
| 10 | Visualization | D3 | v7.9.0 | ISC License (functionally equivalent to | |
| | Visualization | Cytoscape | 3.10.2 | MIT | |
| | Visualization | Vis.js | v9.1.9. | Apache 2.0 / MIT | |
| | Visualization | G6 | 4.8.24 | MIT | |
| 11 | Operational database | Postgres | 16.2 | PostgreSQL license (similar to MIT) | |
| 12 | LLM (more powerful option) | Gemini | 1.5 | | might switch to this LLM from the original one |
| | | | | | |
| | | | | | |
| | | | | | |

| Last Name | First Name | Value | | | | | |
|---|---|---|---|---|---|---|---|
| **Ramesh** | **Sandeepkumar** | 5 | | **5.00** | **OK** | | |
| **Hoffmann** | **Florian** | 5 | | | | | |
| **Rauscher** | **Nikolas** | 5 | | | | | |
| **Ozseker** | **Irem** | 5 | | | | | |
| **Bhesaniya** | **Yash** | 5 | | 0 | No size | | |
| **Fabian Borges** | **Filipe Alexandre** | 5 | | 1 | Trivial size | | |
| **Kotini** | **Kristi** | 5 | | 2 | Small size | | |
| | | | | 3 | Medium size | | |
| | | | | 5 | Large size | | |
| | | | | 8 | Very large size | | |
| | | | | 13 | Too large (size) | | |