Changelog yuuvis® Momentum - Final - 2023 spring

Schlüssel	Zusammenfassung	Beschreibung
AI-1400	New data format for document classifier training	As a data scientist, I would like to have a standardized data format for training of document classifiers, in order to handle training data from different sources more easily. Acceptance criteria: • Each training example is stored in a separate JSON document that contains 　○ document id (property "doc_type") 　○ document class name (property "doc_type") 　○ textual representation of the document (property "string")
AI-1403	Speed optimization for raw document conversion to pickle files	As a Kairos operator, I would like that raw document conversion process runs in parallel, in order to execute it faster and more efficiently. Acceptance criteria: • Document serialization (conversion to pickle files) supports parallel processing. Additional argument *num_workers* specifies how many parallel processes can be used for this purpose. • Dataset creation supports parallel processing. Additional argument *num_workers* specifies how many parallel processes can be used for this purpose. – **Original description with more technical details:** Pre-pipeline step **docs_serialize** (converts raw documents ti pickled docs) and component **create_dataset** (converts pickled documents into pandas dataframe) both sequentially process the data resulting in slow execution times and unused CPU resources. This issue can be mitigated by allowing parallel processing of documents. Lightweight python library joblib can be used for this exact case. Concept for this has been explored in task ~~AI-1387~~ Acceptance criteria: 1. Modify **docs_serialize** in py-Kairos to enable parallel processing. Additional arguments such as *num_workers* should be added. 2. Modify **create_dataset** in entity extraction pipelines to enable parallel processing. Additional arguments such as *num_workers* should be added. 3. Test speed improvements
AI-1404	Speed optimization for featurize components	As a Kairos operator, I would like that extraction of features from training documents runs in parallel, in order to execute it faster and more efficiently. Acceptance criteria: • Extracion of features from preprocessed documents (~~AI-1403~~) supports parallel processing. Additional argument *num_workers* specifies how many parallel processes can be used for this purpose. – **Original ticket description (with technical details)** Pipeline components such as **featurize**, or more precisely, methods for computing **context word** feature sequentially process the dataframe resulting in slow execution times and unused CPU resources. This issue can be mitigated by allowing parallel processing of dataframes groups (pandas *groupby* method). Lightweight python library pandarallel can be used for this exact case. Concept for this has been explored in task ~~AI-1387~~ Acceptance criteria: 1. Modify **featurize** in entity extraction pipelines to enable parallel processing. Additional arguments such as *num_workers* should be added. 2. Test speed improvements

AI-1448	Graph-Neural-Network pipeline for invoice number extraction	**Goal is not to perform major changes on the approach itself (finding/adding new features, making it run faster, etc.) but on code stabilization and tranformint the prototype (with all possible shortcuts) into code that can be shiped and maintained in future releases. All improvements of that kind shall be done in separate tickets (feel free to create the improvement tickets where you find it necessary).** _ As a system operator, I would like to use pipelines for extraction of invoice numbers with high accuracy, in order to provide a high quality extraction results to users of my system. Acceptance criteria: • The GNN (Graph-Neural-Network) prototype is tranformed into production-ready code • The pipeline is documented • Automated tests for the pipeline are implemented
AI-1449	Graph-Neural-Network pipeline for invoice date extraction	**Goal is not to perform major changes on the approach itself (finding/adding new features, making it run faster, etc.) but on code stabilization and tranformint the prototype (with all possible shortcuts) into code that can be shiped and maintained in future releases. All improvements of that kind shall be done in separate tickets (feel free to create the improvement tickets where you find it necessary).** _ As a system operator, I would like to use pipelines for extraction of invoice issuing date with high accuracy, in order to provide a high quality extraction results to users of my system. Acceptance criteria: • The GNN (Graph-Neural-Network) prototype is tranformed into production-ready code • The pipeline is documented • Automated tests for the pipeline are implemented
BUD-2136	Users are supported by an high-contrast theme for better accessibility	As a user, you can select a high-contrast theme for better accessibility. Acceptance Criteria: 1. The color values for contrast are researched and recommended by competent institutions. 2. The Bluebird High contrast theme is tested and good approaches are adopted. 3. The high-contrast theme can be activated in the settings.
BUD-2144	Administrators can configure whether to encode result list exports with BOM	As an administrator, I want to be able to configure whether to encode result list exports with BOM so that users get correct emojis and German Umlauts when opening the csv into excel by clicking the file in the Windows explorer. Acceptance criteria: • It is possible to configure this feature. Default is without BOM. • If configured with BOM excel opens the data encoded with BOM. Add in api-web-prod.yml : csv: writeBOM: true • Works for UTF-8
BUD-2174	The client is based on Angluar 15	As a developer, you can use an updated Angular version. Acceptance criteria: The client is updated to Angular 15.
BUD-2247	A recipient of a task can be a user or a role	As a user assigning a task, you are supported by offering the ability to assign the task to a role and not just to one or more users. Acceptance criteria: 1. In the UI for selecting recipients of a task, you can select whether you are looking for a role or a user to select. 2. It works for identity management with Keycloak, in a second step also for IDM Controller.
BUD-2252	Enhancement of Tenant Management API with roles and Zerberus with new endpoint	As a programmer for BPM, you will be supported in the Tenant Management API with an endpoint for roles. Acceptance Criteria: 1. The Tenant Management API provides roles. 2. Zerberus is extended with an endpoint.

BUD-2262	Create new IDM controller in Zerberus	As a programmer, you are provided with a new IDM Controler in order to have a consistent architecture on the basis of which further extensions are feasible. Acceptance criteria: • New IDM (Identity Management) Controller is created • Endpoints users, whoami and roles from user controller are available in the idm controller • Endpoints users and whoami from user controller are marked as deprecated and kept until the next major release • Technical roles are configurable from api-web-prod.yaml. (Technical roles are used to filter out technical users from get users query.)
BUD-2336	Allow custom result list configuration in main.json	As a programmer, you are allowed to set a default row height for result lists for a custom client. In main.json include custom search configuration like this:"search": { "gridRowHeight": **70**, "limit": **50** } (TBD)

ERA-9055	User can filter tasks by assigned state	As a user, I would like to filter the tasks to those that are assigned to me, those that I can claim, or to see all of them, in order to manage my tasks more effectively.
		
		Acceptance criteria:
		
		• The endpoint for getting tasks has a boolean query parameter "isAssigned"
		  ○ **technical info:** parameter "filter" is removed from GET /tasks/ and replaced by this one, breaking change that has to be documented clearly
		  ○ if parameter is true, user gets only tasks assigned to the user
		  ○ if parameter is false, user gets only tasks visible to user, but not yet assigned to someone (these are the tasks that user can see and can claim them)
		    ■ **technical info**: user is candidate either directly or via role-membership, and no one has claimed the task
		  ○ if parameter is not provided user gets assigned and unassigned tasks

OKTO-5020	Fast deletion with subsequent removal of the documents via a background process	As a developer, I would like to delete documents via the API in such a way that they are no longer accessible to users without having to wait for the entire cycle of deletion to finish in order to achieve a better UX for the user. Acceptance criteria: • it is possible to delete documents in such a way that they are only marked as deleted • Documents deleted in this way are treated as already deleted documents in the search. • the deletion of documents takes place asynchronously via a background process
OKTO-5288	WebHook for deleting metadata can modify metadata and prevent deletion	As a solution developer, I would like to bend the WebHook (dms.request.objects.delete) for deleting documents so that only metadata is changed but the document is not deleted to, for example, prohibit access or implement a recycle bin. Acceptance criteria: • In the WebHook the deletion of documents can be prevented via the WebHook • metadata can be changed when deletion is prevented • The data is adjusted so that it behaves in the further like a metadata update • The extension of the function is documented
OKTO-5383	separate full text request for the content of documents	As a user, I would like to be able to perform full text queries separately for the meta data and the full text of the documents in order to search for information in a more targeted way. Acceptance criteria: • the SQL syntax has been extended so that targeted searches are possible • also in combination with other fields • applies to IDX and content file full text data • tests are available for the feature • the feature is documented
OKTO-5440	Global document type with free assignment of SOTs	As a developer, I would like to provide components that have no further dependency on object types. For this purpose, a basic object type is to be created that is open to all SOTs. Acceptance criteria: • A system type is available to which all SOTs can be added at runtime. • the system type does not have its own properties • the system type is documented • system:Document becomes instantiable • not to be included ○ Configurable document life cycle, currently it can only be determined via WebHooks whether SOTs are allowed.

OKTO-5444	Endpoint for determining the total file size of all documents of a tenant	As a system integrator, I would like to determine the total bytes consumed of the binary files for a client via an endpoint in order to be able to implement quotas.

OKTO-5444 — Endpoint for determining the total file size of all documents of a tenant

As a system integrator, I would like to determine the total bytes consumed of the binary files for a client via an endpoint in order to be able to implement quotas.

Acceptance criteria:

- an endpoint is available via which the data can be provided
- the data is calculated in the background and is available in real time
- No real time data is provided
- Rendition data are not taken into account
- the endpoint is documented

OKTO-5499 — Greedy-parameters for batch-delete

The boolean request parameter "greedy" should be used to control whether as much as possible should be deleted during batch delete.

So far either all objects in the batch are deleted or none. If one cannot be deleted, the whole request fails and no object is deleted (greedy=false, default).
If greedy=true, then all objects are deleted where it is possible.

As a consequence, no error messages can be thrown as before and the response must be reworked.
There are several reasons why an object cannot be deleted, all leading to different error messages with different HttpStatusCodes:

1. no object exists for the ID or it is not visible (HttpStatusCodes =404).
2. the user has no delete permission or retention forbids deletion (HttpStatusCodes =403)
3. the object is a folder which is not empty (HttpStatusCodes =409)

If the deletion is successful, an HttpStatusCodes 200 is returned without a body.

OKTO-5549 — Integrate slide rendition for yMomentum

As a user, I would like to have a slide preview provided for documents to provide a better overview in user interfaces.

Acceptance criteria:

- there is a REST endpoint through which the slide rendition is provided.
- the permission system for the documents is respected.
- the feature is documented.

20 Vorgänge