

Changelog yuuvvis® Momentum - Final - 2023 summer

Schlüssel	Zusammenfassung	Beschreibung
AI-1297	Classification of PDF documents	<p>As a Predict API user, I would like to be able to classify documents presented in format other than fulltext in order to have more usability</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none">• Predict API can accept pdf files instead of a fulltext representation for classification.• Predict API can call pdf-converter to convert the pdf into fulltext for classification model inference
AI-1348	Training pipeline for invoice number extraction is dockerized	<p>As a system operator, I would like to have dockerized classifiers and extractors, in order to be able to manage them more easily (initial installation, update to newer training pipeline, etc.)</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none">• Invoice number extraction pipeline is packed in a docker image<ul style="list-style-type: none">◦ Precondition: Pickling of raw documents has been performed already◦ If not, an error is returned• The training parameters can be set<ul style="list-style-type: none">◦ URL of S3 bucket containing featurized data and credentials for it◦ Number of epochs◦ Batch size◦ Split ratio (training:validation)◦ The name of label to be learned• Training can be executed with these parameters• Training artefacts and metrics are sent to the ML-flow service, to be persisted• The dockerization shall allow to multiple users to start multiple training pipelines, each with its own set of parameters• The training parameters and training pipeline are documented
AI-1415	Tenant administrator can start preprocessing on exported data	<p>As a Kairos user, I would like to start preprocessing of the raw data that I want to use for training later on, in order to convert it to data format that is expected by the Kairos ML-Pipelines.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none">• POST /api/admin/preprocessing• New instance of service that performs preprocessing is spawned and runs on the data• Error is returned if preprocessing is already running on the same data repository
AI-1429	Tenant administrator can see the status of preprocessing operations	<p>As a Kairos user, I would like to know the status of preprocessing operations, in order to be able to manage them, and to know when can I use them for training.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none">• GET /api/admin/preprocessing <ul style="list-style-type: none">◦ user can specify sorting on start time and end time◦ the endpoint supports paging• User gets the running preprocessing operations, including their start time (AI-1415)• User gets the finished preprocessing operations, including their start and end time, as well as their execution summary (AI-1427)• User gets the interrupted preprocessing operations, including their start time and end (interruption) time (AI-1430)
AI-1430	Tenant administrator can interrupt a preprocessing operation	<p>As a Kairos user, I would like to know stop execution of a preprocessing operations in order to better utilize computing resources in the cluster. For instance, this can be used if operation is taking too long, or if it was started erroneously (on a wrong data set).</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none">• POST /api/admin/preprocessing/id • Preprocessing operation is interrupted, preprocessing service execution stops• There is no cleanup of preprocessing results that were saved to the S3 bucket, user shall clean it up according to their needs and wishes

AI-1478	Classification of company names to invoice issuers and invoice receivers	<p>As a service user, I would like to be able to accurately classify company names as invoice issuer or invoice receiver. This will save me time and reduce errors when issuing and receiving invoices.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> Classify company name as invoice issuer or invoice receiver using the machine learning approach ✓ There is a training pipeline for this service ✓ The service is dockerized and provides a Rest-API ✓
AI-1481	Train head for metadata extraction that uses transformer model language	<p>As a kairos user, I would like to extract all relevant information from an invoice accurately in order to be able to manage it more easily.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> Train a head for invoice number extraction Train a head for total amount extraction Train a head for base amount extraction Train a head for invoice number extraction Evaluate trained model on Bär dataset
AI-1501	Predict-Api uses a machine learning classifier for invoice issuer/receiver	<p>As a user of Kairos-AI I would like to differentiate between a company that issued an invoice and company that is receiving the invoice, in order to better support users that are working with large number of invoices.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> Predict-Api is integrated with the machine learning approach for invoice issuer/receiver classification (AI-1478)
AI-1556	Create an endpoint to extract fulltext from pdf	<p>As a kairos user, I would like to be able to classify documents in different input format in order to have more usability.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> PDF-converter supports conversion of pdf to fulltext for classification model inference
AI-1583	Allow GPU usage during inference of dockerized model	<p>As a system operator, I want to enable GPU usage during the inference process of a dockerized machine learning model so that the model can make predictions faster and more efficiently.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> The dockerized model must be configured to detect and utilize GPUs available on the host machine during inference. ✓ The model must support both CPU and GPU modes for inference. ✓ The docker image must include all necessary libraries and dependencies for GPU acceleration. ✓ The model must be tested in both CPU and GPU modes to ensure that predictions are consistent across both configurations. ✓ The performance results shall be documented in a wiki page. ✓
AI-1601	Amount calculation used only in specific profile	<p>Acceptance criteria:</p> <p>Amount calculations will be performed only if specific profile is activated</p>
AI-1602	Remove unused request/response types	<p>There are some request/response types used before that are still supported in code but will never be used in that format in the future. They should be removed</p>

BUD-360	Label for an user or an object by Web-API Gateway GET endpoints	As a programmer, you want to be able to resolve the display name of a user by Web-API Gateway GET endpoints so that I can present them instead of the ID or formatting it in the frontend.
		Acceptance criteria:
		<ul style="list-style-type: none"> • All endpoints that return OrganizationObject should now return a modified one with property "displayName" which contains formatted label. • User display name formatting should be configurable in api-web yaml configuration • Every user id resolver in bpm and dms should use the same formatting
		<p>Doku:</p> <p>There is a new property available in zerberus configuration. The property named "api-web.user-label-templates" contains templates for creating display name for Users when resolving user ids or returning users as OrganizationObjects. These templates are formatted by finding "%{property}" and resolving it with one of the user fields (available options are "id", "email", "username", "firstname" and "lastname"). Since it could be possible that one of these fields is empty api-web will try to resolve each template in the list in order until one of them is succesfull (none of the required properties are blank). If none are resolved, id will be set as display name.</p>
		<p>Example:</p> <p>api-web.user-label-templates:</p> <ul style="list-style-type: none"> - "%{lastname}, %{firstname} (%{username})" - "%{firstname} (%{username})" - "%{lastname} (%{username})" - "%{username}"
BUD-1961	Modularization of yM architect and refactoring (Schema management, forms modeling, configuration)	As a developer, you will be supported in your work with the ym Architect by being able to use individual modules. Development with the yM Architect becomes easier, and faster because more efficiency is achieved.
		Acceptance criteria:
		<ul style="list-style-type: none"> • There is a new architect GIT repository 'admin-apps' with the monorepo that has integrated the architect with these states <ul style="list-style-type: none"> ◦ Schema management ◦ forms modeling, ◦ configuration • Some technical debts are resolved, e.g. unnecessary complexity in multiple states • The new architect is deployed on the Team-CI-system
BUD-2202	System responsables are supported by a viewer service docker image that is reduced in size (depedency: for yuuviz RAD)	As a system responsible, I want to be supported by a viewer service docker image that is reduced in size so that it does not need too many resources.
		Acceptance criteria:
		<ul style="list-style-type: none"> • The size of the service is reduced by removing libraries (e.g. node modules) that are not needed, with that possible security issues reduced as well.
BUD-2453	Adaptation for log out flow in tenant management API	As a user you get a comfortable log out flow in the client.
		<p>Prerequisite:</p> <p>The authentication service is adapted in such a way that the TM API can make an adjustment without causing bugs.</p>
		There is a system in which this adapted authentication service is available and can be used for testing with the new TM API.
		Acceptance criteria:
		<ul style="list-style-type: none"> • the TM API is adapted for the new log out flow. • The reference client uses this new log out flow. • This modification is documented.
BUD-2455	Keycloak themes are adopted for the docker image with version 21	The keycloak themes are adopted for the docker image with version 21.
BUD-2456	Improved basis of one-click indexing	The one-click indexing feature can be used as a plug-in in the yM Reference client and in all custom clients. For better performance, the code of the KCIM feature for the one-click indexing is fundamentally redesigned.

BUD-2491	Search metadata only	As a user, I am supported by a search that I use to search for metadata only.
		Acceptance Criteria:
		<ul style="list-style-type: none"> • It is easy for users to see whether they are doing a full-text search "on everything." Or whether the search relates to metadata. • The search in metadata has no hits if the string searched for is not present in the metadata. • The search in metadata has hits only if the searched string is present in the metadata. • Searching only in metadata is at least as performant as full-text search. • The search in metadata only uses the capability introduced by Octopus to separate the full text into binary and metadata. See OKTO-5383 separate full text request for the content of documents
ERA-8477	Process instances can be deleted	As workflow administrator, I would like to delete process instances, in order to remove those that are no longer needed.
		Acceptance criteria:
		<ul style="list-style-type: none"> • Process instances can be (multi)selected • The selected instances can be deleted • If some of delete operations fail, the batch-delete continues running (attempts to delete other processes from the list) • Outcome of operation is shown to the user <ul style="list-style-type: none"> ◦ In case of an error while deleting, user is notified about the affected process ◦ In case that all were successfully deleted, user is notified that all were successfully deleted • After deletion operation (successful or not), current page is reloaded and no processes are selected
OKTO-3213	Delete binary content of a document	As a user I want to delete the binary content of documents. The document should only contain meta data afterwards.
		Acceptance criteria:
		<ul style="list-style-type: none"> • there exists in endpoint to delete the binary content • the document is searchable via the meta data • the document is no longer searchable via the text of the binary content • there is an audit entry • the function is documented in the official WIKI
OKTO-5236	Validation when changing schemas	As a developer, I would like to check what effects a change to the schema has on the dataset so that I can better deal with updates to the schema.
		Acceptance criteria:
		<ul style="list-style-type: none"> • the following things should be checked: <ul style="list-style-type: none"> ◦ Removal of an object type from the schema. ◦ Removal of a property type from the schema. ◦ Removal of a SOT from an object type (but SOT remains in the schema). ◦ Removal of a PropertyType (PropertyReference) from an object type (PropertyType remains in the schema). ◦ Change the Cardinality of a PropertyType from Multi to Single.
OKTO-5534	Optional 'abac' section in internal JWT	As a system integrator I would like to have the possibility to extend the internal JWT with a project specific section. This section contains extended user information (e.g. all mailing lists he is member of).
		Acceptance Criteria:
		<ul style="list-style-type: none"> • The internal JWT may have an 'abac' node with extended user information. • The 'abac' node contains one or more named string lists. • Structure and size of the extended JWT are validated.
OKTO-5536	'abac' Attributes in Roleset Conditions	As a system integrator or tenant administrator, I would like to be able to define roles whose conditions contain 'abac' attributes .
		Acceptance Criteria:
		<ul style="list-style-type: none"> • The conditions in Rolesets can use the project specific 'abac' information from the internal JWT. • Example: @abac.mailGroups • the function is documented

OKTO-5540	Calling the background erase without wait parameter	<p>As a developer, I also want to delete faster while not waiting for the delete job to complete to achieve a better UX.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> • the parameter waitForSearchConsistency can be used during deletion • the parameter is documented
OKTO-5546	Optimisation of memory consumption during text extraction of Office documents	<p>As a user, I would like to be able to process large office documents in such a way that the texts can be extracted as much as possible so that I can work with complete documents.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> • large Excel and PowerPoint documents can be processed by the CA in such a way that the CA does not crash. • the feature is tested and documented
OKTO-5562	Native support of the Azure Blobstore	<p>As a system integrator, I would like to connect an Azure Blobstore natively for better system integrity and stability.</p> <p>Acceptance Criteria:</p> <ul style="list-style-type: none"> • an Azure blobstore can be configured as a binary store • Data can be stored in the Blobstorage and queried • there is a possibility to use the configured retentions of the blob storages • the feature is documented
OKTO-5569	Thumbnails for e-mails	<p>As a user, I would like thumbnails to be displayed for emails so I can get a better preview in results lists.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> • thumbnails can be created for emails • the thumbnails are stored in the RenditionCache • the thumbnails for emails can be fetched via the Rendition endpoint of the API • if no thumbnail exists, it will be created during the request via the API
OKTO-5573	Slide rendition for .pbm and .jng files	<p>As a user, I would like the pbm and jng formats to also be supported in slide creation to support better format coverage.</p> <p>Acceptance criteria:</p> <ul style="list-style-type: none"> • slides are created for the formats in question • the feature is documented
OKTO-5624	Metric endpoint for the number of documents of a tenant	<p>As a system integrator, I would like to know how many documents a tenant has stored in order to enable billing or to ensure that no tenants are deleted that still contain data. The standard search is not suitable for this purpose, since the authorization system takes effect and a user would otherwise have to be set up for this query, who must always have full authorization to access all documents.</p> <p>Acceptance Criteria:</p> <ul style="list-style-type: none"> • a new metrics endpoint is available that provides the number of all documents in the system for a tenant • When determining the number, the authorization system is not taken into account • the feature is documented
OKTO-5625	Support for the new Office 365 formats in mime type recognition	<p>As a user of the cloud-Office365 integration, after editing and automatically saving the documents, I would like to have them automatically made available again in the Office365 integration so that I can work without interruption.</p> <p>The current Office365 version, unlike the previous version, has made changes to the file format. These mean that the specific document types are no longer recognized, but only that it is an Office format. In Office integration, this causes workflow issues.</p> <p>Acceptance Criteria:</p> <ul style="list-style-type: none"> • When uploading Office365 files, it tries to recognize which types they are. • If only a global type is recognized, the file extension of the transferred file name is also used for analysis. • The change is documented. • The change is communicated.

