Key Point Analysis – Parameters Summary

green parameters are changed frequently
yellow ones can be used but with cautious
red parameters should probably be avoided

Supported run_params (used when starting an analysis job):

Parameter	Allowed Values	Default	Explanation
arg_min_len	Integer	3	Filter shorter sentences (by number of tokens)
arg_max_len	Integer	36	Filter longer sentences (by number of tokens)
arg_relative_aq_threshold	Float in (0,1)	1.0	Filter arguments having Argument Quality (AQ) score below this precentile
arg_min_aq	Float in (0,1)	0.0	Filter arguments having AQ score below this threshold
mapping_policy	“STRICT”, ”NORMAL”, “LOOSE"	“NORMAL”	Policy for determining if an argument is matched to a key point: for “strict”, only pairs with very high matching scores will be considered matched, leading to a higher precision and a lower coverage, and vice versa for loose. Can’t be used along with “mapping_threshold”.
kp_granularity	“FINE”, “NORMAL”, “COARSE”	“NORMAL”	Used for key point selection: choose “FINE” for more fine grained key points, and “COARSE” for more distinct key points. Can’t be used along with “clustering_threshold”.
clustering_threshold	Float in (0,1)	0.999	Used for key points selection: choose higher values for more fine-grained key points, and lower for distinct key points. Can’t be used along with “kp_granularity”
matching_threshold	Float in (0,1)	clustering_threshold	Threshold for matching arguments to key points inside the selection algorithm
merging_threshold	Float in (0,1)	clustering_threshold	Threshold for merging similar candidates inside the selection algorithm
mapping_threshold	Float in (0,1)	0.999	The matching threshold, scores above are considered a match. A higher threshold leads to a higher precision and a lower coverage. Can’t be used along with “mapping_policy”
sentence_to_multiple_kps	Boolean	False	Can a sentence be matched to more than one key point (above threshold)? When false the sentence is matched the key point with the highest match score
arg_subset_frac	Float in (0,1)	internal algorithm	The fraction of unique sentences to be used in the first stage of inference
max_sentences_for_first_stage	Integer	5000	If arg_subset_frac is not set, the internal algorithm can be limited by this value

n_top_kps	Integer	internal algorithm	Number of top key points to generate. All arguments are remapped to these key points.
non-inference	Boolean	False	When set to True, will throw an exception if new inference is required
kp_extractor_two_stages	Boolean	True	When set to False, will disable two stage mechanism and run one stage
kp_relative_aq_threshold	Float in (0,1)	1	Arguments having AQ score below this precentile will not be selected as key point candidates
kp_min_aq	Float in (0,1)	0.3	Arguments having AQ score below this threshold will not be selected as key point candidates
invalid_kps_comment_ids	String list	Empty list	A list of comment_ids who's sentences should not be selected as key point candidates
kps_first_sentences_only	Boolean	False	Should only the first sentence of a comment be selected as a key point candidate (reducing the missing context problem)
kp_pronoun_filter	Boolean	False	Should sentences that start with pronoun be filtered and not be selected as a key point candidate (reducing the missing context problem) prinoun list: ['i', 'it', 'this', 'its', 'these', 'they', 'he', 'she', 'that']
filter_kps_similar_to_motion	String	None	If there are key points that are too similar to the motion (the topic), it is possible to pass the motion in this parameters and they will be filtered out
similarity_to_motion_threshold	Float in (0,1)	merging_threshold	The threshold above which a key point is considered too similar to the motion
use_stage1_sentences_in_stage2	Boolean	False	Set to True if only the key points are needed and there is no need to map all sentences to them. The result will contain the key points and the sentences that are used in stage 1.
stances_to_run	List of strings	None	A list of stances. Sentences with stance score above stances_threshold will be used. Other sentences will be filtered out. Only available if stance was calculated when comments where uploaded.
stances_threshold	Float in (0,1)	0.5	Must be supplied when stances_to_run is used Only available if stance was calculated when comments where uploaded.
kp_min_len	Integer	3	Minimal length of key points (by number of tokens)
kp_max_len	Integer	8	Maximal length of key points (by number of tokens)

use_kp_quality	Boolean	True	When set to true, use the Key Point Quality (KPQ) classifier for key points selection.
kp_min_kp_quality	Float in (0,1)	0.9	Arguments having KPQ score below this threshold will not be selected as key point candidates * Relevant only when use_kp_quality is true.
sort_candidates_by_kp_quality	Boolean	True	If set to True, first filter out candidates whose KPQ score below the threshold, and then sort according to KPQ score and select top candidates. If set to False, sorting is done according to AQ score and not KPQ score. * Relevant only when use_kp_quality is true.
use_sentence_embeddings	Boolean	True	If set to False, will not use sentence embeddings and will use the slower matching service instead.
enable_sentence_embeddings_on_the_fly	Boolean	True	If set to False, will require that all sentence embeddings will be precalculated and stored in the DB (using do_sentence_embeddings in domain_params). * Relevant only when use_sentence_embeddings is true.
sentence_embedding_filter_factor_stage1	Float	0.2	For each sentence, this is the ratio of top scored key points (using sentence embeddings) that KPA will also use the matching model over, in the first stage (key points selection). * Relevant only when use_sentence_embeddings is true.
sentence_embedding_n_top_kps_stage1	Int	5	Minimum number of top scored key points (using sentence embeddings) that KPA will also use the matching model over, in the first stage (key points extraction). * Relevant only when use_sentence_embeddings is true.
sentence_embedding_filter_factor_stage2	Float	0.1	For each sentence, this is the ratio of top scored key points (using sentence embeddings) that KPA will also use the matching model over, in the second stage (sentence to key point matching). * Relevant only when use_sentence_embeddings is true.
sentence_embedding_n_top_kps_stage2	Int	5	Minimum number of top scored key points (using sentence embeddings) that KPA will also use the matching model over, in the second stage (sentence to key point matching). * Relevant only when use_sentence_embeddings is true.
keypoints **	List of strings	Empty list	When keypoints are provided the service matches the sentences to the given keypoints

			instead of extracting them automatically. This enables a human-in-the-loop scenario, where the automatically extracted key points are reviewed by the user, and the sentences are then remapped to the revised keypoints.
keypoints_by_job_id **	String	None	It is also possible to use key points from a previous job by suppling the job_id in this param.
keypoints_by_job_ids **	List of strings	None	Similar to *keypoints_by_job_id* except you can pass multiple job_ids and all key points from these jobs will be used.
keypoint_candidates **	List of strings	Empty list	When keypoint_candidates are provided, the service matches the sentences to the provided keypoint_candidates but also automatically extracts new ones, if exist.
keypoint_candidates_by_job_id **	String	None	It is also possible to provide keypoint_candidates by taking key points from a previous job by suppling the job_id in this param.
keypoint_candidates_by_job_ids **	List of strings	None	Similar to *keypoint_candidates_by_job_id* except you can pass multiple job_ids and all key points from these jobs will be used.

** Only one of these parameters should be used

Supported domain_params (used when creating a new domain):

Parameter	**Allowed Values**	**Default**	**Explained**
dont_split	Boolean	False	When set to True, the comments uploaded to the domain will not be cleaned and not be split into sentences.
dc	String		a dominant-concept to use as a topic for the argument_quality service
motion	String		a motion to use as a topic for the argument_quality service (either use motion or dc, not both)
do_stance_analysis	Boolean	False	When set to true, stance is calculated for all sentences

do_kp_quality	Boolean	True	When set to true, keypoint quality is calculated for all sentences with num tokens <= kp_quality_max_tokens
do_sentence_embeddings	Boolean	False	When set to True, stores the sentence embeddings for each sentence (storage space consuming)
kp_quality_max_tokens	Int	15	Will calculate the keypoint quality score for sentences with <= tokens
rabbitmq_infer_queue_name	String		Sets a new queue/model for match-score calculation
rabbitmq_sbert_queue_name	String		Sets a new queue/model for sentence embeddings
rabbitmq_aq_queue_name	String		Sets a new queue/model for argument quality calculation
rabbitmq_kpq_queue_name	String		Sets a new queue/model for key point quality calculation
rabbitmq_stance_queue_name	String		Sets a new queue/model for stance calculation
insert_method	SKIP, OVERRIDE	SKIP	Changes the behavior when uploading a comment that was already uploaded (according to the comment_id). SKIP: will not save the new one. OVERRIDE with replace the old one.