

Partial Extractor List

Scientific

Domain	Tool	Description	Repositories
Hydrology	Advection Diffusion	Solve a general advection-dispersion equation.	
	Chemical Mean Age	Determine the mean age of chemical constituents with inputs of chemical dynamics.	
	Document Tables Extractor (ncaa.nlp.wordtables)	Extract tables from documents.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/WordTablesExtractor
	GDAL (ncaa.geo.shpExtractor)		https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-geo/browse
	GDAL (ncaa.geo.tiffExtractor)		https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-geo/browse
	GDAL (ncaa.image.geotiff)		https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-geotiff/browse
	Historical River Extractor (ncaa.cv.river)	Extract the river networks from the ancient hand-drawing maps and compare them with current river networks	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/river
	Normalized Difference Vegetation Index (ncaa.arccgis.landsat7mosaic)		https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-bd-cz/browse/ndviextractor
	River Chi Index	Identify the river dynamics in a river basin and evaluate human activities' influences through Chi index in the streams.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-bd-cz/browse/chi-analysis
	River Sinuosity	Study the maturity and equilibrium conditions of a stream through the sinuosity index.	
	Soil Moisture Change	Determine role of hydraulic redistribution in AZ (riparian site / upland site) by studying soil moisture change throughout different seasons.	
	Species Classifier	SAM based Species Classification from Hyperspectral data, Hyperspectral Indices, NDVI, SAVI, MSAVI, etc.	
	TerEx (ncaa.arccgis.floodplain)	Identify the flat polygons and the heights inside a river valley.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-bd-cz/browse/terex_floodplain
	Topographic Depressions	Identify topographic depressions (TDs) and their distribution on landscape (Number, location, area, volume of TDs).	https://github.com/HydroComplexity/TDI
	Valley Safety Zones	Estimate submerging areas and water depths under extreme floods and map the safety zones in a river valley.	
	Vegetation Indices	Calculating vegetation indices such as NDVI and Surface Temperature from Landsat 7 and 8 satellite data.	
Ecology	PlantCV (terra.plantcv)	Extract plant height, area, and color distribution from photographs.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-plantcv/browse
Civil & Environmental Engineering	Body of Water Detector (ncaa.image.ponddetect)	Land coverage, extract locations of bodies of water from satellite data.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-maps/browse/feature_detection
	GI Identification		
	Human Preference Score (ncaa.image.humanpref)	Assign a model derived human preference score to a given image of an urban environment.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-maps/browse/humanpref
	Route Greenness (ncaa.xml.greenindexroute)	Derive the green index of a city route.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-maps/browse/greenroute
	Social Media GI Preferences		
	Stanford CoreNLP - Sentiment (ncaa.nlp.SNLPSentiment)	Assign a sentiment score to a piece of text.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/SNLP/SNLPSentimentExtractor
Social Science & Humanities	TUV Triaxus		
	Bertillon Card Cell Extractor (ncaa.image.dmp)	Extract table cells from a Bertillon Card.	https://opensource.ncaa.illinois.edu/bitbucket/projects/DEBOD/repos/extractors-debod

	Census Form Cell Extractor (census-section-segmentor)	Extract table cells from a 1930s Census form.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/census
	Handwritten Decimals Extractor (ncsa.image.sphog.debod)	Extract handwritten decimal values from an image.	https://opensource.ncsa.illinois.edu/bitbucket/projects/DEBOD/repos/extractors-handwrittendecimals
	Killed Photos (ncsa.image.iarp_remove_circle)	Identify depression era photos "killed" by Farm Security Administration director Roy Stryker, indicated by a hole punch in the image.	https://opensource.ncsa.illinois.edu/bitbucket/projects/IARP/repos/image_fetcher/browse/extractors/remove_circle
	Mean Grey (ncsa.cv.meangrey)	Mean grey values of black and white photos.	https://opensource.ncsa.illinois.edu/bitbucket/projects/IARP/repos/image_fetcher/browse/extractors/mean_grey
	Movie Slice (ncsa.movieslice)		https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-movieslice/browse
	Person Detector (person-detector)	Extract locations of people in an image.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-person-detector
	Person Tracker (ncsa.person-tracker)	Extract locations and paths of people moving in videos.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-person-tracking
	Video Analytics Toolbox (ncsa.cinemetrics_batch)	Extract shot descriptors from videos.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-lsva-integrated/browse
Biology, Genomics, Medicine	FSL (mri2mesh)		https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-mri/browse/mri2mesh
	Glomeruli (ncsa.msc.diagnosis)	Extract glomeruli from Kidney biopsy images.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-msc/browse/Kidney

General

Domain	Tool	Type	Description	Repositories
	CMU Sphinx (ncsa.audio.speech2text)	Extract or	Audio recognition to extract text for speech within audio.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-core/browse/audio/speech2text
	DBPedia (ncsa.dbpedia)	Extract or	Find and define named entities within the given text.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-dbpedia/browse
	ImageMagick (ncsa.image.metadata)	Extract or	Pull available EXIF image metadata from a given image.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-core/browse/image/metadata
	OpenCV - Faces (ncsa.cv.faces)	Extract or	Find faces in an image and return their locations.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-opencv/extractors-opencv-faces
	OpenCV - Eyes (ncsa.cv.eyes)	Extract or	Find eyes in an image and return their locations.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-opencv/extractors-opencv-eyes
	OpenCV - Closeups (ncsa.cv.closeups)	Extract or	Determine whether an image is a closeup of a person or not.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-opencv/extractors-opencv-closeups
	OpenCV - Profiles (ncsa.cv.profiles)	Extract or	Find human face profiles in an image and return their locations.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-opencv/extractors-opencv-profiles
	Langid (ncsa.nlp.simplelanguage)	Extract or	Identify the language of the given text.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/SimpleLanguage
	NLTK - Summary (ncsa.nlp.simplesummary)	Extract or	Summarize a body of text.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/SimpleSummary
	Siegfried (siegfried)	Extract or	Extract information about a given file relevant to identifying its type and validating its format.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-siegfried/browse
	Stanford CoreNLP (ncsa.nlp.SNLP)	Extract or	Natural Language Process extractions such as parts of speech, named entities, language, etc.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/SNLP/SNLPExtractor
	Tesseract (ncsa.image.ocr)	Extract or	Object Character Recognition (OCR) to extract text from images containing text.	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-tesseract
	Tika (ncsa.nlp.tika)	Extract or	Document extractions such as language identification, ...	https://opensource.ncsa.illinois.edu/bitbucket/projects/CATS/repos/extractors-nlp/browse/tika

General	Versus - Color Distribution (ncaa.versus.image)	Extract or	Generate a distribution of color values within an image to be used for comparing how similar two images are.	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-versus/browse
	VLFeat (ncaa.image.caltech101)	Extract or	Classify images as to whether they contain objects from the Caltech101 dataset (e.g. people, airplanes, motorcycles, cougars, ...).	https://opensource.ncaa.illinois.edu/bitbucket/projects/CATS/repos/extractors-cv/browse/extractors-vlfeat

Older [Extractors](#)