

# A tool for recommending keywords with more live and more attention

## Supplementary Material

### WordNet and CSO comparison

Down below, the Table S1 summarizes a list of common words included in WordNet and CSO ontology. The algorithm is limited to explore at only a distance of two neighbours from the original word.

Original word	WordNet refinement	Computer Science Ontology (CSO) refinement
argumentation	discussion	computer_programming_languages
accelerometer	measuring instrument	sensors
administrative_data_processing	data processing	database
anonymity	anonymity	privacy
artificial_intelligence	robotics	artificial_intelligence
authentication	validation	security_of_data
binoculars	binoculars	binocular
biometrics	eugenics	access_control
blog	diary	internet
broadcasting	radio	communication_channels_%28information_theory%29
buffer_storage	fund	bandwidth
cad	scoundrel	computer_aided
channel_capacity	channel_capacity	communication_channels_%28information_theory%29
chaotic_attractor	attractor	cryptography
classification_system	edition	control_system
clustering	meet	clustering
computer_programming	computer_programming	computer_programming
computer_science	plan	software
computer_virus	computer_virus	security_of_data
computer-aided_design	software	computer_aided
computerized_tomography	tomography	medical_imaging
contract	flex	contract
control_system	servo	control_system
cosmic_microwave_background	cosmic_microwave_background	polarimeter
cryptography	cryptography	cryptography
cryptology	psychology	cryptography

data_mining	data_processing	clustering
demodulation		communication_channels_
	reception	information_theory
e-commerce	e-commerce	internet
edutainment	entertainment	e-learning
electrical_energy	AC	electrical_energy
electromagnetism	acoustics	electromagnetic
electronic_mail	spam	internet
elementary_education	pedagogy	e-learning
ergonomics	technology	human_computer_interaction
facial_expression	frown	facial_expression
file_system	file_system	computer_science
flash_memory	flash_memory	embedded_system
gateway	gateway	routing_protocols
handover	release	mobile_telecommunication_systems
holmium	lu	rare_earth
industrial_management	technology	information_technology
knowledge_base	realm	artificial_intelligence
lexical_database	lexical_database	artificial_intelligence
logic_gate	logic_gate	hardware
logic_programming	programming	computer_programming_languages
machine_translation	robotics	artificial_intelligence
magnetic_storage	magnetic_storage	microprocessor_chips
memory_access	memory_access	memory_access
mobile_phone	cell	sensors
network_architecture	spec	computer_network
neural_network	rf	neural_network
nuclear_physics	crystallography	nuclear_physics
object-oriented_programming	hack	java
optical_fiber	loofa	sensors
outage	breakdown	outage
owl	raptor	semantic_web
phase_modulation	fm	sensors
privacy	privacy	privacy
programming_language	prolog	programming_language
purchasing	purchasing	internet
reconstruction	makeover	reconstruction
relational_database	relational_database	database
remote_control	device	robotics
robotics	robotics	robots
search_engine	program	internet
sip	ingestion	internet_protocol
source_code	source_code	software_engineering
speckle	speckle	radar
spin	revolution	spin
spline	remove	computer-aided_design
surveillance_system	surveillance_system	cryptography
tactics	tactics	software_design
target_language	language	machine_translation
telecommunication_equipment	television	sensors
teleconferencing	discussion	telecommunication_services
transmission_control_protocol	http	internet_protocol
user_interface	CLI	sensors

validation	validation	validation
verification	checksum	verification
virtual_storage	memory	database
visual_communication	graphics	image_coding
web_page	web_page	internet
white_noise	impediment	white_noise
wireless	wireless	cellular

Table S1: WordNet vs CSO refinement. The list of original words is included in both ontologies (WordNet and CSO).

## Survey structure

### Question 1

**Title:** HUBBLE: an optical link management system for dense wavelength division multiplexing networks

#### Abstract

Timely detection of Dense Wavelength Division Multiplexing (DWDM) link quality and service performance problems of fiber deployment are important and critical for telecommunication operators. In this paper, we propose a new methodology for network fault detection inside optical transmission systems deployed in a real-operator environment and present the working principles of the system. Our new calculation methodology is used for joint fiber and DWDM link quality evaluation inside the proposed High-level Unified BackBone Link Examiner (HUBBLE) platform. At the end of the paper, we also detail some of the benefits, challenges, and opportunities of automation in DWDM networks using the proposed HUBBLE platform.

Initial set	Refined set
Dense wavelength division multiplexing	D.W.D.M
service provider	service provider
fault management	network services
link quality	link quality

### Question 2

**Title:** Feature Completion for Occluded Person Re-Identification

## Abstract

Occluded person re-identification (Re-ID) is a challenging problem due to the destruction of occluders. Most existing methods focus on visible human body parts through some prior information. However, when complementary occlusions occur, features in occluded regions can interfere with matching, which affects performance severely. In this paper, different from most previous works that discard the occluded region, we propose a Feature Completion Transformer (FCFormer) to implicitly complement the semantic information of occluded parts in the feature space. Specifically, Occlusion Instance Augmentation (OIA) is proposed to simulate real and diverse occlusion situations on the holistic image. These augmented images not only enrich the amount of occlusion samples in the training set, but also form pairs with the holistic images. Subsequently, a dual-stream architecture with a shared encoder is proposed to learn paired discriminative features from pairs of inputs. Without additional semantic information, an occluded-holistic feature sample-label pair can be automatically created. Then, Feature Completion Decoder (FCD) is designed to complement the features of occluded regions by using learnable tokens to aggregate possible information from self-generated occluded features. Finally, we propose the Cross Hard Triplet (CHT) loss to further bridge the gap between complementing features and extracting features under the same ID. In addition, Feature Completion Consistency (FC2) loss is introduced to help the generated completion feature distribution to be closer to the real holistic feature distribution. Extensive experiments over five challenging datasets demonstrate that the proposed FCFormer achieves superior performance and outperforms the state-of-the-art methods by significant margins on occluded datasets.

Initial set	Refined set
Training	Training
Task analysis	Task analysis
Feature extraction	Feature extraction
Image retrieval	Image retrieval
Optimization	Local convergence
Semantics	Linguistics terminology
Person re-identification	Person re-identification
Data mining	virtual_learning

### Question 3

**Title:** Robust Adaptive Fault-Tolerant Control of Multiagent Systems With Uncertain Nonidentical Dynamics and Undetectable Actuation Failures

#### Abstract

This paper studies the distributed consensus problem of multiagent systems (MASs) in the presence of nonidentical unknown nonlinear dynamics and undetectable actuation failures. Of particular interest is the development of a robust adaptive fault-tolerant consensus protocol capable of compensating uncertain dynamics/disturbances and time-varying yet unpredictable actuation failures simultaneously. By introducing the virtual parameter estimation error into the artfully chosen Lyapunov function, the consensus problem is solved with a robust adaptive fault-tolerant control scheme based upon local (neighboring) agent state information. It is shown that the proposed method is user friendly in that there is no need for detail dynamic information of the agent or costly detection/diagnosis of the actuation faults in control design and implementation, resulting in a structurally simple and computationally inexpensive solution for the leaderless consensus problem of MAS. Simulation results illustrate and verify the benefits and effectiveness of the proposed scheme.

Initial set	Refined set
Fault tolerant control	Fault tolerant control
Adaptative control	Mobile agent
Multiagent systems	Speed control

### Question 4

**Title:** Mobile Robot Obstacle Avoidance Based on Neural Network with a Standardization Technique

#### Abstract

Reactive algorithm in an unknown environment is very useful to deal with dynamic obstacles that may change unexpectantly and quickly because the workspace is dynamic in real-life applications, and this work is focusing on the dynamic and unknown environment by online updating data in each step toward a specific goal; sensing and avoiding the obstacles coming across its way toward the target by training to take the corrective action for every possible offset is one of the most challenging problems in the field of robotics. This problem is solved by proposing an Artificial Intelligence System (AIS), which works on the behaviour of Intelligent Autonomous Vehicles (IAVs) like humans in recognition, learning, decision making, and action. First, the use of the AIS and some navigation methods based on Artificial Neural Networks

(ANNs) to training datasets provided high Mean Square Error (MSE) from training on MATLAB Simulink tool. Standardization techniques were used to improve the performance of results from the training network on MATLAB Simulink. When it comes to knowledge-based systems, ANNs can be well adapted in an appropriate form. The adaption is related to the learning capacity since the network can consider and respond to new constraints and data related to the external environment.

Initial set	Refined set
Obstacle avoidance	Autonomous vehicles
Simulation-based learning	Simulation-based learning
Neural networks	Radial basis function
Autonomous mobile robots	Autonomous robots

## Question 5

**Title:** Deep temporal motion descriptor (DTMD) for human action recognition

### Abstract

Spatiotemporal features have significant importance in human action recognition, as they provide the actor's shape and motion characteristics specific to each action class. This paper presents a new deep spatiotemporal human action representation, the deep temporal motion descriptor (DTMD), which shares the attributes of holistic and deep learned features. To generate the DTMD descriptor, the actor's silhouettes are gathered into single motion templates by applying motion history images. These motion templates capture the spatiotemporal movements of the actor and compactly represent the human actions using a single 2D template. Then deep convolutional neural networks are used to compute discriminative deep features from motion history templates to produce the DTMD. Later, DTMD is used for learning a model to recognize human actions using a softmax classifier. The advantage of DTMD are that DTMD is automatically learned from videos and contains higher-dimensional discriminative spatiotemporal representations as compared to handcrafted features; DTMD reduces the computational complexity of human activity recognition as all the video frames are compactly represented as a single motion template; and DTMD works effectively for single and multiview action recognition. We conducted experiments on three challenging datasets: MuHAVI-Uncut, iXMAS, and IAVID-1. The experimental findings reveal that DTMD outperforms previous methods and achieves the highest action prediction rate on the MuHAVI-Uncut dataset.

Initial set	Refined set
Human activity recognition	Human activity recognition
Deep convolutional neural network	Deep convolutional neural network

Motion history images	Motion history images
Deep temporal motion descriptor	Deep temporal motion descriptor
Computer vision	Linear motor

## Question 6

**Title:** Well placement optimization using metaheuristic bat algorithm

### Abstract

The design of an optimal field development and production management is a complicating task because of influencing various factors on decision-making process. Typical factors include number and type of wells, well locations, production constraints, economic factors like capital expenditure, operating costs, and oil sale price. The situation is further complicated due to the uncertainty associated with various effective engineering and geological parameters.

In this study, three meta-heuristics algorithms of genetic algorithm (GA), particle swarm optimization (PSO) and bat inspired algorithm (BA) are used for optimal determination of six production well locations. Net present value (NPV) is used as an objective function in optimization process. PUNQ-S3 benchmark model is simulated in MATLAB environment in order to search the entire complex reservoir during optimization. Next, the effectiveness of algorithms will be compared in terms of convergence rate and NPV improvement over iterations.

The simulation results show that the BA is superior since it reduces the number of functional evaluations and thus improving the computational efficiency. In addition, the BA provides better NPV improvement over PSO and GA. The results indicate that the BA increases NPV by 7.5% and 21.7% over PSO and GA respectively.

Initial set	Refined set
Fractional calculus	Fractals
Bat algorithm	Bat algorithm
Levy flight	Nonlinear equations
Non-parametric statistical tests	Non-parametric statistical tests

## Question 7

**Title:** An Iterated Multi-stage Selection Hyper-heuristic**Abstract**

There is a growing interest towards the design of reusable general purpose search methods that are applicable to different problems instead of tailored solutions to a single particular problem. Hyper-heuristics have emerged as such high level methods that explore the space formed by a set of heuristics (move operators) or heuristic components for solving computationally hard problems. A selection hyper-heuristic mixes and controls a predefined set of low level heuristics with the goal of improving an initially generated solution by choosing and applying an appropriate heuristic to a solution in hand and deciding whether to accept or reject the new solution at each step under an iterative framework. Designing an adaptive control mechanism for the heuristic selection and combining it with a suitable acceptance method is a major challenge, because both components can influence the overall performance of a selection hyper-heuristic. In this study, we describe a novel iterated multi-stage hyper-heuristic approach which cycles through two interacting hyper-heuristics and operates based on the principle that not all low level heuristics for a problem domain would be useful at any point of the search process. The empirical results on a hyper-heuristic benchmark indicate the success of the proposed selection hyper-heuristic across six problem domains beating the state-of-the-art approach.

Initial set	Refined set
Hyperheuristics	Cognitive systems
Simultaneous multithreadings	Simultaneous multi-threading
Resource partitioning	Resource partitioning
Fuzzy partition	Optimization techniques

**Question 8**

**Title:** An alternative method of biomedical signal transmission through the GSM voice channel

**Abstract**

In this work, a new solution for online and accurate biomedical data transmission is presented. For this purpose, a global system for mobile (GSM) communication voice channel is, for the first time, used as a communication link between the patient and healthcare provider. Biomedical signals are converted into speech-like signals before being transferred over a GSM



voice channel. On the receiver side, speech-like symbols are stored in a symbols bank, and constructed using random stochastic signals. On the receiver end, the index of the symbol with the most similarity to the received signal is selected as the identified sample. This method enables us to communicate with an accuracy of 99.8% at a transfer rate of 110 samples per second and signal-to-noise ratio (SNR) of 10. By utilizing a GSM voice channel, any voice channel, such as a cell phone, can be used for data transmission. The transmitted signal is encoded; therefore, the connection is secured. GSM technology has benefits such as availability, reliability, and robustness. Additionally, GSM can be used as a backup or service for transmitting vital physiological signals in emergency situations (e.g. in an ambulance). This technology can also be used to transmit other physiological signals as well as nonphysiological generic data.

Initial set	Refined set
Electrocardiography	Mathematics in medicine
GSM	GSM
Telemedicine	Telemedicine
Voice	Voice
Speech codecs	Speech codecs

## Question 9

**Title:** Characterizing Generalized Rate-Distortion Performance of Video Coding : An Eigen Analysis Approach

### Abstract

Rate-distortion (RD) theory is at the heart of lossy data compression. Here we aim to model the generalized RD (GRD) trade-off between the visual quality of a compressed video and its encoding profiles (e.g., bitrate and spatial resolution). We first define the theoretical functional space  $W$  of the GRD function by analyzing its mathematical properties. We show that  $W$  is a convex set in a Hilbert space, inspiring a computational model of the GRD function, and a method of estimating model parameters from sparse measurements. To demonstrate the feasibility of our idea, we collect a large-scale database of real-world GRD functions, which turn out to live in a low-dimensional subspace of  $W$ . Combining the GRD reconstruction framework and the learned low-dimensional space, we create a low-parameter eigen GRD method to accurately estimate the GRD function of a source video content from only a few queries. Experimental results on the database show that the learned GRD method significantly outperforms state-of-the-art empirical RD estimation methods both in accuracy and efficiency. Last, we demonstrate the promise of the proposed model in video codec comparison.

Initial set	Refined set
Rate-distortion function	Rate-distortion function
Video quality assessment	Video quality
Quadratic programming	Quadratic programming

## Question 10

**Title:** WONDER: Weighted One-shot Distributed Ridge Regression in High Dimensions

### Abstract

In many areas, practitioners need to analyze large datasets that challenge conventional single-machine computing. To scale up data analysis, distributed and parallel computing approaches are increasingly needed.

Here we study a fundamental and highly important problem in this area: How to do ridge regression in a distributed computing environment? Ridge regression is an extremely popular method for supervised learning, and has several optimality properties, thus it is important to study. We study one-shot methods that construct weighted combinations of ridge regression estimators computed on each machine. By analyzing the mean squared error in a high dimensional random-effects model where each predictor has a small effect, we discover several new phenomena.

1. Infinite-worker limit: The distributed estimator works well for very large numbers of machines, a phenomenon we call "infinite-worker limit".
2. Optimal weights: The optimal weights for combining local estimators sum to more than unity, due to the downward bias of ridge. Thus, all averaging methods are suboptimal.

We also propose a new Weighted ONE-shot DistributEd Ridge regression (WONDER) algorithm. We test WONDER in simulation studies and using the Million Song Dataset as an example. There it can save at least 100x in computation time, while nearly preserving test accuracy.

Initial set	Refined set
Distributed learning	Learning methods
High-dimensional statistics	High-dimensional statistics
Ridge regression	Optimization techniques
Random matrix theory	Random variable

