Annotation

This annotation task aims to compare the relatedness of arXiv classes used to classify preprints uploaded on arXiv.org. Each record consists of a triple: the anchor class and two classes (A, B). The label should identify which class (A or B) is more related to the given anchor class. Use “A” or “B” to indicate which class is semantically more related to the given “Anchor” class. If no relation between the two classes and the anchor class exists, or if it is not possible to differentiate which is more related, use “0” (try to avoid this case). The classes are defined based on the following descriptions.

Class descriptions

Artificial Intelligence	Covers all areas of AI except Vision, Robotics, Machine Learning, Multiagent Systems, and Computation and Language (Natural Language Processing), which have separate subject areas. In particular, includes Expert Systems, Theorem Proving (although this may overlap with Logic in Computer Science), Knowledge Representation, Planning, and Uncertainty in AI. Roughly includes material in ACM Subject Classes I.2.0[Artificial Intelligence], I.2.1[Applications and Expert Systems], I.2.3[Deduction and Theorem Proving], I.2.4[Knowledge Representation Formalisms and Methods], I.2.8[Problem Solving, Control Methods, and Search], and I.2.11[Distributed Artificial Intelligence].
Hardware Architecture	Covers systems organization and hardware architecture. Roughly includes material in ACM Subject Classes C.0, C.1[Processor architectures], and C.5[Computer system implementation] .
Computational Complexity	Covers models of computation, complexity classes, structural complexity, complexity tradeoffs, upper and lower bounds. Roughly includes material in ACM Subject Classes F.1 [Computation by abstract devices], F.2.3 [Tradeoffs among complexity measures], and F.4.3 [Formal languages], although some material in formal languages may be more appropriate for Logic in Computer Science. Some material in F.2.1[Numerical Algorithms and Problems] and F.2.2[Non Numerical Algorithms and Problems], may also be appropriate here, but is more likely to have Data Structures and Algorithms as the primary subject area.
Computational Geometry	Roughly includes material in ACM Subject Classes I.3.5[Computational Geometry and Object Modeling] and F.2.2[Non Numerical Algorithms and Problems].
Computation and Language	Covers natural language processing. Roughly includes material in ACM Subject Class I.2.7[Discourse, Language generation, Language models, Language parsing and understanding, Machine translation, Speech recognition and synthesis, Text analysis]. Note that work on artificial languages (programming languages, logics, formal systems) that does not explicitly address natural-language issues broadly construed (natural-language processing, computational linguistics, speech, text retrieval, etc.) is not appropriate for this area.
Cryptography	Covers all areas of cryptography and security including authentication, public key

and Security	cryptosystems, proof-carrying code, etc. Roughly includes material in ACM Subject Classes D.4.6[Access controls, Authentication, Cryptographic controls, Information flow controls, Invasive software] and E.3[Data encryption].
Computer Vision and Pattern Recognition	Covers image processing, computer vision, pattern recognition, and scene understanding. Roughly includes material in ACM Subject Classes I.2.10[Vision and scene understanding], I.4[Image processing and computer vision], and I.5[Pattern recognition].
Databases	Covers database management, datamining, and data processing. Roughly includes material in ACM Subject Classes E.2[Data storage representations], E.5[Files], H.0[Information Systems], H.2[Database Management], and J.1[Administrative data processing].
Distributed, Parallel, and Cluster Computing	Covers fault-tolerance, distributed algorithms, stability, parallel computation, and cluster computing. Roughly includes material in ACM Subject Classes C.1.2[Multiple Data Stream Architectures (Multiprocessors)], C.1.4[Parallel Architectures], C.2.4[Distributed Systems], D.1.3[Concurrent Programming], D.4.5[Reliability], D.4.7[Batch processing systems, Distributed Systems, Hierarchical design, Interactive systems, Real-time and embedded systems], E.1[Data structures].
Digital Libraries	Covers all aspects of the digital library design and document and text creation. Note that there will be some overlap with Information Retrieval (which is a separate subject area). Roughly includes material in ACM Subject Classes H.3.5[Online information services], H.3.6[Library Automation], H.3.7[Digital Libraries], I.7[Document and text processing].
Discrete Mathematics	Covers combinatorics, graph theory, applications of probability. Roughly includes material in ACM Subject Classes G.2[Combinatorics, Graph Theory] and G.3[Probability and statistics].
Formal Languages and Automata Theory	Covers automata theory, formal language theory, grammars, and combinatorics on words. This roughly corresponds to ACM Subject Classes F.1.1[Models of Computation], and F.4.3[Formal Languages]. Papers dealing with computational complexity should go to **Computational Complexity**; papers dealing with logic should go to **Logic in Computer Science**.
Graphics	Covers all aspects of computer graphics. Roughly includes material in all of ACM Subject Class I.3[Computer Graphics], except that I.3.5[Computational Geometry and Object Modeling] is likely to have Computational Geometry as the primary subject area.
Computer Science and Game Theory	Covers all theoretical and applied aspects at the intersection of computer science and game theory, including work in mechanism design, learning in games (which may overlap with Learning), foundations of agent modeling in games (which may overlap with Multiagent systems), coordination, specification and formal methods for non-cooperative computational environments. The area also deals with applications of game theory to areas such as electronic commerce.
Human-Compu	Covers human factors, user interfaces, and collaborative computing. Roughly includes

ter Interaction	material in ACM Subject Classes H.1.2[User/Machine Systems] and all of H.5[Information interfaces and presentation], except for H.5.1[Multimedia Information Systems], which is more likely to have Multimedia as the primary subject area.
Information Retrieval	Covers indexing, dictionaries, retrieval, content and analysis. Roughly includes material in ACM Subject Classes H.3.0[Information storage and retrieval], H.3.1[Content analysis and indexing], H.3.2[Information storage], H.3.3[Information search and retrieval], and H.3.4[Systems and Software].
Information Theory	Covers theoretical and experimental aspects of information theory and coding. Includes material in ACM Subject Class E.4[Data compaction and compression, Error control codes, Formal models of communication] and intersects with H.1.1[General systems theory, information theory, value of information].
Machine Learning	Papers on all aspects of machine learning research (supervised, unsupervised, reinforcement learning, bandit problems, and so on) including also robustness, explanation, fairness, and methodology. **Machine Learning** is also an appropriate primary category for applications of machine learning methods.
Logic in Computer Science	Covers all aspects of logic in computer science, including finite model theory, logics of programs, modal logic, and program verification. Programming language semantics should have Programming Languages as the primary subject area. Roughly includes material in ACM Subject Classes D.2.4[Software/Program Verification], F.3.1[Specifying and verifying and reasoning about programs], F.4.0[Mathematical logic and languages], F.4.1[Mathematical Logic], and F.4.2[Grammars and other rewriting systems]; some material in F.4.3[formal languages] may also be appropriate here, although Computational Complexity is typically the more appropriate subject area.
Multiagent Systems	Covers multiagent systems, distributed artificial intelligence, intelligent agents, coordinated interactions. and practical applications. Roughly covers ACM Subject Class I.2.11[Distributed Artificial Intelligence].
Multimedia	Roughly includes material in ACM Subject Class H.5.1[Multimedia Information Systems].
Mathematical Software	Roughly includes material in ACM Subject Class G.4[Mathematical software].
Numerical Analysis	Numerical algorithms for problems in analysis and algebra, scientific computation. Roughly includes material in ACM Subject Class G.1[Numerical analysis].
Networking and Internet Architecture	Covers all aspects of computer communication networks, including network architecture and design, network protocols, and internetwork standards (like TCP/IP). Also includes topics, such as web caching, that are directly relevant to Internet architecture and performance. Roughly includes all of ACM Subject Class C.2[Computer-communication networks, Network architecture and design, network protocol, network operations, local and wide-area network, internetworking] except C.2.4[Distributed Systems], which is more likely to have Distributed, Parallel, and Cluster Computing as the primary subject area.

Operating Systems	Roughly includes material in ACM Subject Classes D.4.1[Process management], D.4.2[Storage memory], D.4.3[File system management], D.4.4[Communications management], D.4.5[Reliability], D.4.7[Batch processing systems, distributed systems, hierarchical systems, interactive systems, real-time systems and embedded systems], and D.4.9[Systems programs and utilities].
Programming Languages	Covers programming language semantics, language features, programming approaches (such as object-oriented programming, functional programming, logic programming). Also includes material on compilers oriented towards programming languages; other material on compilers may be more appropriate in Architecture (AR). Roughly includes material in ACM Subject Classes D.1[Programming techniques] and D.3[Programming languages].
Robotics	Roughly includes material in ACM Subject Class I.2.9[Autonomous vehicles, commercial robots and applications, kinematics and dynamics, manipulators, operator interfaces, propelling mechanisms, sensors, workcell organization].
Symbolic Computation	Roughly includes material in ACM Subject Class I.1[Expressions and their representation, algorithms, languages and systems].
Sound	Covers all aspects of computing with sound, and sound as an information channel. Includes models of sound, analysis and synthesis, audio user interfaces, sonification of data, computer music, and sound signal processing. Includes ACM Subject Class H.5.5[Sound and music computing], and intersects with H.1.2[User/Machine Systems], H.5.1[Multimedia information systems], H.5.2[User interfaces], I.2.7[Natural language processing], I.5.4[Hypertext/Hypermedia], I.6.3[Simulation and Modelling applications], J.5[Arts and humanities], K.4.2[Social Issues].
Software Engineering	Covers design tools, software metrics, testing and debugging, programming environments, etc. Roughly includes material in all of ACM Subject Classes D.2[Requirements/Specifications, Design tools and techniques, coding tools and techniques, testing and debugging, programming environments, Distribution maintenance and enhancement, Software architectures, Interoperability, Reusable software], except that D.2.4[Program verification] should probably have Logics in Computer Science as the primary subject area.
Systems and Control	This section includes theoretical and experimental research covering all facets of automatic control systems. The section is focused on methods of control system analysis and design using tools of modeling, simulation and optimization. Specific areas of research include nonlinear, distributed, adaptive, stochastic and robust control in addition to hybrid and discrete event systems. Application areas include automotive and aerospace control systems, network control, biological systems, multiagent and cooperative control, robotics, reinforcement learning, sensor networks, control of cyber-physical and energy-related systems, and control of computing systems.