ISTHMUS: History as an Answer Key for Cross-Field Retrieval

DarkSigma, Inc.

2 September 2026

Abstract

Cross-field retrieval has no test collection, because the relevance judgments are ordinarily written by the group that built the system under evaluation. We take them from documented history instead: fifty cases in which a researcher solved a problem by importing a mechanism from a second field, each recorded by a published paper. Each problem is restated in the asking field's vocabulary as of the period before the transfer, and that restatement is the only input any arm receives.

Given those problems, a commercial literature-search service whose index is larger than ours by roughly three orders of magnitude returns the asking field's own literature in 91 to 96 per cent of results at every rank band, and nothing else at all in 37 of the 50 cases. It reaches the field the historical case names in 3. Our engine, on an index three orders of magnitude smaller, reaches it in 21 ($p = 3.6 \times 10^{-5}$).

A third arm sends our own domain-neutral restatement of each problem to the same index. It leaves the asking field about as often as our engine does, 190 adjudicated papers against 192, and reaches the field that solved the problem in 2 of 50 cases. Leaving the asking field and arriving in the field that solved it are separate capabilities, and rewriting the query supplies only the first.

Nine hundred returned papers were adjudicated blind by two instruments agreeing on 93 percent of labels. Of the out-of-field work, 61 of 192, 5 of 18 and 34 of 190 satisfy the relevance criteria, which score eight historical source papers 8 of 8 and eleven mismatched pairings 0 of 11.

1 What is being measured

Mechanisms transfer between fields, and the transfers are documented: simulated annealing from the statistical mechanics of cooling solids into combinatorial optimization (Kirkpatrick et al., 1983), compressed sensing from sparse signal recovery into magnetic resonance imaging (Lustig et al., 2007), chaos control from nonlinear dynamics into cardiology (Garfinkel et al., 1992). Appendix B lists fifty such cases. Retrieving the source work from a statement of the problem, before the transfer has been made, is the task measured here.

A researcher states a problem in the vocabulary of their own field, and the system is required to return work from a field the researcher does not read which carries a mechanism applicable to that problem. Evaluating this task is difficult because no test collection exists for it. The relevant document lies in a literature the assessor

does not know in advance, so relevance judgments are ordinarily produced by the same group that built the system under evaluation.

We obtain relevance judgments from documented history instead. In each of our fifty cases a researcher solved a problem by importing a mechanism from a second field, and a published paper records the borrowing. Garfinkel and colleagues stabilized an arrhythmic rabbit heart in 1992 using the chaos-control method Ott, Grebogi and Yorke had published in 1990; the connection was made by hand and the 1992 paper is its record. The judgment therefore predates the system, and a miss is scored against a connection a person made without it.

We report four measures of increasing strictness: whether the system returned any work from outside the asking field, whether it reached the specific field the historical case names, what fraction of the out-of-field work satisfies the relevance criteria of §8.3, and whether the system returns the specific paper the transfer was built on. §11 reports all four.

2 Related work

Representing a paper by its purpose and its mechanism, so that documents from different fields can be matched on structure, is not new. Hope, Chan, Kittur and Shahaf built purpose and mechanism vectors for product descriptions and showed that retrieval over them surfaces analogies a vocabulary-based retriever does not.[1] Chan, Chang, Hope, Shahaf and Kittur extended the representation to research papers in SOLVENT, where annotators mark the background, purpose, mechanism and findings of each abstract and a model built from those annotations retrieves analogous work across fields.[2] Analogies retrieved through SOLVENT were judged useful by domain experts and outperformed a vocabulary-matching baseline.

Kang, Qian, Hope, Shahaf, Chan and Kittur later removed the annotation step, building an end-to-end analogical search engine over papers and reporting that the fully automated system reached accuracy comparable to their human-in-the-loop one.[3] Automatic construction of this representation is therefore established, and the present work does not claim it.

What differs here is the evaluation. Theirs is expert judgment of retrieved analogies, collected after retrieval by the group that built the system, against problems those scientists brought. Ours is taken from documented history, so the judgments exist before the system runs and were recorded by people with no stake in it.

Scoring a system against discoveries that already happened is the standing practice in literature-based discovery, beginning with Swanson's reconstruction of the fish-oil and Raynaud's syndrome connection from pre-1986 literature.[4] The field's usual form is time slicing: a corpus is cut at a date, candidates are generated from the earlier part, and later co-occurrence in the literature counts as confirmation. Moreau surveys what that costs, arguing the field is evaluated on too few discoveries and on a target, co-occurrence, that is mostly noise rather than discovery.[5] The move made here is not new; the unit is what differs. A case is a documented transfer in which a published paper states what was borrowed and from which field, so the target is a specific paper rather than a term pair, and the fifty cases span disciplines rather than sitting inside one biomedical index.

On design, the benchmark follows standard test-collection practice: a fixed topic set, a shared rank cutoff, pooled results, blind assessment, and reported assessor agreement. BEIR is the closest analogue in shape, a heterogeneous zero-shot retrieval benchmark assembled from existing collections.[6] BEIR inherits annotations made for the underlying collections; we take ours from the historical record.

1. T. Hope, J. Chan, A. Kittur, D. Shahaf. Accelerating Innovation Through Analogy Mining. *KDD*, 2017. arXiv:1706.05585.
2. J. Chan, J. C. Chang, T. Hope, D. Shahaf, A. Kittur. SOLVENT: A Mixed Initiative System for Finding Analogies between Research Papers. *Proc. ACM Hum.-Comput. Interact.* 2(CSCW), Article 31, 2018.
3. H. B. Kang, X. Qian, T. Hope, D. Shahaf, J. Chan, A. Kittur. Augmenting Scientific Creativity with an Analogical Search Engine. *ACM Transactions on Computer-Human Interaction*, 2022. arXiv:2205.15476.
4. D. R. Swanson. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine* 30(1):7–18, 1986. doi:10.1353/pbm.1986.0087.
5. E. Moreau. Literature-based discovery: addressing the issue of the subpar evaluation methodology. *Bioinformatics* 39(2):btad090, 2023. doi:10.1093/bioinformatics/btad090.
6. N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, I. Gurevych. BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. *NeurIPS Datasets and Benchmarks*, 2021. arXiv:2104.08663.

3 What a case is

A case is one historical transfer recorded as ten fields. The case is the unit of the benchmark and the denominator of every case-level rate below. Fifty were written by hand.

Table 1. The fields of a case and the role each plays.

Field	Contents and use
`id`	Stable slug, the join key across every evidence file.
`fieldA`	The asking field, whose problem it is. Its literature is what the researcher already reads, so returning it does not count.
`fieldB`	The source field the mechanism came from. One field that historically worked. The case makes no claim that it was the only one.
`problem`	The query, in the first person, in field A's vocabulary, as of the period before the transfer. Supplied unchanged to every arm.
`targetTerms`	Five terms native to field B and absent from field A. Used for the leak audit in §4, for the corpus-coverage probe in §7, and as a term-overlap diagnostic printed beside the run that takes no part in any score.
`bridgePaper`	The publication that made the transfer. It postdates the problem statement by construction, since the problem is written as of the period before the transfer and this paper is the transfer. Excluded from scoring on every arm.
`doi`	Identifier for that publication, so a reader can confirm the case is real.
`lane`	Coarse tag for the kind of crossing, used only for grouping.
`borderlineFields`	Boolean, true on seven cases where the two fields are close. Reported as context and never used to exclude a case. See §5.1.
`borrowingQuote`	The sentence in the bridge paper that states the borrowing. No scorer reads it.

3.1 A case in full

Case `chaos-control-cardiac-arrhythmia`, reproduced from the case file without paraphrase.

> **Asking field.** Cardiology and cardiac electrophysiology.
>
> **Source field.** Nonlinear dynamics and chaos control, specifically the Ott, Grebogi and Yorke method of stabilising unstable periodic orbits.
>
> **Problem, as supplied to every arm.** "In a rabbit heart made arrhythmic by a drug, the intervals between beats become irregular instead of settling into a steady rhythm. The two available responses both act on the whole tissue at once: load it with more drug, or deliver a single large shock. We would like instead to restore a regular rhythm using only small, ordinary pacing stimuli of the kind a pacemaker already delivers, and we do not know whether stimuli that small can work at all, or when each one would have to be delivered."
>
> **Terms native to the source field.** unstable periodic orbit; OGY chaos control; Poincaré section and return map; stable and unstable manifolds; saddle fixed point.
>
> **Bridge paper.** Garfinkel, Spano, Ditto & Weiss, "Controlling Cardiac Chaos," *Science* 257(5074):1230–1235, 1992.

The problem statement contains no orbit, no attractor, no map and no manifold, which is the condition the leak audit enforces across all fifty cases.

4 Building the case set, and the leak audit

We admitted a transfer to the case set on two conditions: that a published paper states the borrowing, so the judgment is on record and not ours; and that the borrowing is identifiable well enough to name the source field without ambiguity. We did not select for difficulty, and we did not run any arm before the set was closed.

Those conditions select for consequence, because a transfer that mattered is a transfer someone wrote up in a venue that recorded it. Twenty-five of the fifty bridge papers appeared in *Science* (13), *Nature* (7), *PNAS* (2), *Cell*, *The Lancet* and *Physical Review*. Among them are simulated annealing into combinatorial optimization, compressed sensing into MRI, neural-network representations of potential-energy surfaces, non-equilibrium thermodynamics into generative models, liquid-liquid phase separation into germ-cell biology, spin-glass theory into protein folding, and knot theory into DNA recombination. Several founded the subfield they landed in. The set is therefore a sample of transfers that succeeded and were noticed, not a sample of attempted transfers.

The bridge papers span 1958 to 2022 with a median of 2006, and the fifty cases divide across six directions of travel: physics into biology (10), chemistry into distant fields (9), computation into the physical sciences (9), biology into engineering (9), medicine and industrial practice (8), and geoscience with industrial process engineering (5). Each case was written by hand. We did not automate the work: a case is only useful if a person has confirmed the historical link and then removed the vocabulary that would disclose it.

Table 2. Leak rules, the check applied to each, and the outcome across all fifty cases.

Rule	Check	Outcome
No term native to the source field appears in the problem statement.	Every `targetTerm`, split on slashes and parentheses, matched case-insensitively against the normalized problem text.	**Fifty of fifty pass**, zero occurrences. The rule tests each problem statement against five source-field terms recorded in the same case file by the same author, so it detects a leak through those five terms and cannot detect one phrased around them.
No term coined by the bridge paper appears in the problem statement.	Every word longer than seven characters in the bridge paper's title, matched against the problem text.	**Not usable as written.** It flags 22 to 24 cases depending on how the threshold is read, and every flag is a field-A word an honest field-A problem statement must contain: granules, asthmatic, concrete, coughing, software, handover. The bridge paper is written by field A, so its title necessarily shares field A's vocabulary.
The asking field and the source field are not the same field under a different name.	Recorded by hand in `borderlineFields`, true on seven cases.	Reported as context. See §5.1 for why it is not an exclusion.

5 Typing the transfer

Table 3. Transfer types and their distribution across the fifty cases.

Type	What crossed	Cases
MECHANISM	A physical or biological effect. Mussel foot protein chemistry becoming a universal coating; gecko toe adhesion explained by van der Waals forces.	23
METHOD	A mathematical or algorithmic technique. Chaos control into cardiology; knot theory into DNA recombination; optimal transport into single-cell biology.	16
INSTRUMENT	A measurement or fabrication apparatus. Photolithography into DNA arrays; coda-wave ultrasound into concrete.	5
PRACTICE	An organizational routine. Six Sigma into hospitals; evidence-based medicine into software engineering.	6

5.1 Adjacent fields are counted

Seven cases are marked `borderlineFields` because the two fields are close. They are counted in every figure in this document, and the flag is printed beside them.

Cardiology and nonlinear dynamics are adjacent fields, and they are separate ones. A case counts as a crossing when a publication was required to connect the two. Garfinkel and colleagues obtained a paper in *Science* from that connection in 1992, and the existence of that paper is the evidence the gap was real. Excluding a case because its labels look close would discard a crossing that a published paper records.

The failure that flag guards against is real, and it sits one level down, at the individual result. A DNA-topology paper carrying a venue label of Mathematics, returned for a DNA problem, is a label artefact. Adjudicating the returned paper's content catches it, as §9 describes, and no case is dropped on that account.

6 The three arms

We give every arm the identical problem statement and nothing else. No arm receives the source field, the target terms, the bridge paper or the transfer type.

6.1 Arm A, our engine

The problem statement is submitted to our own retrieval system and the single ranked list it returns is what gets scored. No narrower sub-list from the response is read. Generation downstream of retrieval is not exercised.

Two conditions are asserted before a case may be scored: the returned list must be non-empty, and the response must confirm the query was served by the intended configuration and not a degraded fallback. A case that cannot satisfy both after four attempts aborts the run. The entry point, the parameters, the response field read and the configuration values are pinned in the harness. A description of them is not enough to reproduce the arm exactly; the harness is available on request.

Two settings decide whether the pool is drawn from one retrieval lane or two. Every case record carries the settings it ran under, and the fifty cases were also re-run under the deployed configuration, where the pool comes entirely from the abstraction lane. Re-running all fifty cases under the deployed configuration, adjudicated once over a sample drawn by the same rule, gives 48 of 50 cases leaving the asking field against 49, and 22 to 24 reaching the field the case named against 21. The difference between the two configurations is smaller than the difference between the arms. The figures reported below come from the run adjudicated end to end by two instruments with its controls.

6.2 Arm B, the comparison

A commercial literature-search service, general purpose and widely used, whose index is larger than ours by roughly three orders of magnitude. It is not named here. Naming it would turn a measurement of two retrieval strategies into a comparison of two products, and the strategies are what this benchmark is about. No setting on either arm was tuned against the other, and the service was queried through its documented interface with nothing changed from its defaults except the number of results requested.

6.3 Arm C, our query rewrite on their corpus

Arms A and B differ in the corpus and in the query. When arm A returns out-of-field work and arm B does not, that result cannot say which of the two produced it.

Arm C holds the corpus fixed and changes only the query. Each problem statement is passed through the same abstraction operator arm A uses, which restates it in domain-neutral functional language, and the restatement is sent to the endpoint arm B was sent the literal problem. Same endpoint, same page size, same cut-off, same exclusions.

The operator samples, so its output is not deterministic. Every restatement was written to disk before any search ran, so the exact query behind each result is on record and rescoring uses the same text.

The restatements were audited for the same leak the problem statements are audited for in §4. Across the 251 target terms recorded in the case file, none appears in the corresponding neutral query, by exact phrase or under a relaxed test requiring only that every word of the term appear somewhere in it. The operator also preserves the shape of what it is given: the sentence count is unchanged in 50 of 50 cases, the sentence order and relative lengths are preserved, and the function-word skeleton has a median longest-common-subsequence ratio of 0.70 against the original. It rewrites the vocabulary of a problem statement without restructuring it.

6.4 One cut-off, every arm

Every arm is asked for 100 results and scored to a shared cut-off of 20, held in a single constant used by every scorer and every report. Results beyond the cut-off are retained as diagnostics and take no part in the comparison. Requesting a hundred results costs no more than requesting twenty on any arm, so there is no cost argument for scoring the arms to different depths.

7 The corpus-coverage control

A miss has two possible causes that call for different responses: the ranking failed, or the source field is absent from the index. A single hit rate confounds them, and the resulting number largely measures corpus size. We therefore probe the corpus for each case's target terms before searching it, by exact substring match over titles and abstracts. The probe is a database query and uses no model. The corpus held literature matching the target terms for 46 of the 50 cases. The four misses are recorded and are retained in the denominators of every figure below, where they count against the arm that missed them. Scored over the 46 covered cases instead, arm A returns out-of-field work in 45, reaches the field the case named in 19 and returns at least one paper satisfying the criteria in 27; arm B gives 12, 3 and 3, and arm C 35, 2 and 14. The probe separates a ranking failure from an absent literature; the ordering of the arms is unchanged either way.

8 Scoring

8.1 Why scoring reads the papers

A returned paper could be counted a hit whenever one of the case's target terms appears in its title or abstract. That measure cannot separate a mechanism from a homograph. Viscosity, sought for droplet physics, matches a paper on crumb rubber in asphalt. Vascular, sought for coatings modelled on plant vasculature, matches *Receptogenesis in a Vascularized Robotic Embodiment*. Entrainment, sought for the turbulent gas cloud of a cough, matches cloud condensation nuclei over the equatorial Pacific. The failure operates on every arm, so the comparison it produces is uninterpretable in either direction. Term overlap is printed beside the run as a diagnostic and takes no part in any score.

A returned paper's field classification could instead be compared against the case's source field, and that fails more seriously. A paper on tangle analysis of DNA-protein complexes carries a venue-derived label of Mathematics, which under label scoring reads as mathematics surfacing for a genetics problem and therefore as exactly the behavior the benchmark exists to detect. It is a DNA-topology paper written for molecular biologists. A coarse label describes a venue, and the papers that matter most here are the ones whose venue and whose audience disagree. Field labels take no part in any score.

8.2 Whose problem the paper solves

Each returned paper is placed in one of three communities. A paper native to field B, whose authors work on their own field's problems, is the historical match. A paper native to field A is the researcher's own literature. A paper native to neither is a third field, which is a common outcome: an engine can leave the asking field without arriving in the field the case names.

Adjudication is blind, and the set given to the adjudicator carries the item, the case, the two candidate field descriptions, the problem, the title and the abstract, and nothing else. The arm that produced each item is held in a separate key file the adjudicator never reads. Field parity across items and arm balance within 15 percent are asserted before the set is written.

8.3 Relevance criteria

Membership of a community does not establish relevance. We therefore apply a second set of criteria, designed so that two assessors reading the same abstract reach the same verdict. Each criterion is a question about the content of the paper; none asks whether the idea is promising.

The criteria are applied through a single template: *the authors obtained this effect by this cause under these conditions, and the reader requires it under different conditions.* A paper for which neither the effect nor the cause can be named from the abstract fails on that ground. Where the template can be instantiated, we ask whether the effect persists under the reader's conditions.

Two outcomes disqualify a paper: the effect ceases to exist under the reader's conditions, whether by reversal, collapse, or domination by a competing term; or the two settings do not differ by degree and no mechanism connects them. All other outcomes are treated as passes, including the case where the effect persists but realis-

ing it requires engineering the paper does not supply. Gecko tape took a decade, and polydopamine arrived twenty-six years after the mussel-protein work it came from. A check that treated either as broken would reject the papers this benchmark is built on.

This criterion applies to physical effects and not to procedures: a procedure has no physical effect that can cease to exist, and an algorithm executes independently of the reader's operating conditions. Where the transferred item is a method, an algorithm, a control scheme or an organizational routine, we substitute a precondition criterion, and ask whether the reader's setting satisfies the assumptions the procedure requires. Compressed sensing requires the signal to be sparse in some representation. A Kalman filter requires the quantity of interest to be observable from the measurements. A method that stabilizes an unstable periodic orbit requires such an orbit to exist. A precondition that could not fail excludes nothing, so where the only thing linking paper and problem is a general capability, nothing is being carried and the paper does not pass. Work needed to satisfy a precondition is still a pass, on the same grounds as before.

Two more checks follow. The first asks whether the reader's field was already working this way before and independently of this paper, which excludes the paper and everything downstream of it, because a successful transfer eventually looks routine and the source paper is often what made it so. And a grade on the evidence: demonstrated in an operating setting, demonstrated in a laboratory, or simulated only. The grade is recorded and no check reads it, since a simulation is weaker evidence for the same effect and the checks decide whether the effect is there.

What is deliberately not asked is whether the reader can see what to do with it. Deciding a gap is worth crossing, and crossing it, requires the reader's own domain knowledge. This benchmark measures whether the material arrives.

The test carries a control at each end. Eight historical source papers, drawn from the cases and known to have worked, are put through it. So are eleven pairings assembled to be wrong, each taking a real source paper and offering it against a problem it has nothing to do with. Both sets of scores are reported in §11.

9 The adjudication instrument

A language model asked to classify papers returns confident, internally consistent and repeatable answers whether or not it is measuring anything. We therefore treat the adjudicator as an instrument and audit it before scoring any result with it.

We give the instrument no menu of labels. It first names, from the paper alone, the research community that would cite it. It then asks two independent graded questions, each seeing only that community name and one of the two candidate fields. Neither question knows the other exists, so there is no ordering to reverse. A paper inside field B alone is the historical match, inside field A alone belongs to the asking field's own literature, and inside both or neither is assigned to a third field.

Table 4. The four pre-registered checks. Any one of them disqualifies the instrument.

Check	What it catches	Bar	Result
Known answer	The bridge paper applies field B from outside, so it must not come back as the historical match. Nor may a paper from an unrelated case.	80%	**16/16**
Repeat	The same item three times. Disagreement means the score is noise.	80%	**7/8**
Order	The two independent questions asked in the reverse order. A moved answer means they are not independent.	80%	**8/8**
Blinding	Items must carry identical fields and the arms must be balanced, so nothing identifies the producing engine.	none	**pass**

Asking whether a community is *part of* a field has a direction, and the case file writes fields narrowly, so "cell and developmental biology" is not part of "Cell and developmental biology (germ-cell specification in *C. elegans*)". The instrument uses a graded relation, which has no direction, and only the strictest grade places a paper inside a field. That costs recall and protects precision.

9.1 Two independent passes

Four checks establish that an instrument is stable, order-invariant, blind and correct on a fixed set assembled in advance. None establishes that it is correct on the 900 items it is about to score, and an instrument that is consistently and impartially wrong passes all four. We therefore adjudicate all 900 items twice, using two instruments that differ in how the question is put, and report the agreement between them.

The second instrument sees the abstract as well as the title and applies one test: at an ordinary conference of the asking field, would this paper be a normal talk by a member of that community, or an invited outsider. It is given a worked example of each of the two ways that question goes wrong. A crystal structure prediction paper shown to a crystallographer is the reader's own literature even though the search algorithm came from evolutionary computation. A materials paper on synthetic dry adhesives shown to a lizard biomechanist belongs to the materials community, even though it is about geckos.

The two agree on 837 of 900 individual labels, which is 93 percent. The disagreements run almost entirely one way: 56 of the 63 are papers the second instrument places in the reader's own literature and the first does not.

Table 5. The case-level field figures under both instruments. The first is reported throughout this document and the second is the sensitivity bound.

	A, first	A, second	B, first	B, second	C, first	C, second
Cases leaving the asking field	49	48	13	4	39	36
Cases reaching the field the case named	21	21	3	2	2	2

Five of the six figures move by three cases or fewer, and question 2 for arm A is 21 under both. The exception is arm B's cases-leaving figure, which moves from 13 to 4 because the second instrument is the stricter of the two about what counts as the reader's own literature, and arm B returns more of that literature than either other arm. The first instrument's 13 is reported throughout; the second widens our margin over arm B on both questions and over arm C on question 1, and leaves it unchanged at 21 against 2 on question 2. Both label sets are published.

No figure in this document is produced by the model family the engine under test runs on.

10 Run integrity

All 50 cases produced results on all three arms. No case required a retry and none aborted. We assert that the corpus is reachable before the run starts and abort on failure. An empty record is never written, so a case that returns nothing means the engine returned nothing. The comparison service was queried once per case per arm, with no retries.

10.1 What is examined, and what is not

The three arms returned 14,929 results in total. We discard ranks beyond the shared cutoff, leaving 2,967 eligible results, and adjudicate 900 of them, 300 per arm, or 30.3 percent of the eligible population. Selection is a seeded hash over the eligible list, so it is random with respect to content and reproducible, and no scorer takes any part in choosing it.

Adjudication reads the title and, where one can be found, the abstract. Abstracts are looked up by title from public sources for every arm through one pipeline; the commercial endpoint returns its own abstracts and they are discarded, because their formatting would identify the arm and break the blind. Coverage is 83 percent on arm A, 75 percent on arm B and 83 percent on arm C, a spread of 8 points against a bar of 15. The remaining 177 items are judged from the title alone.

The case-level counts below are lower bounds: sampling can hide a case's only out-of-field result and cannot invent one. The per-paper rates are two-sided estimates over the sample and can move in either direction. The remaining 2,067 eligible results were not examined.

11 Result

The comparison arm sets the level the other results are read against. Given a problem written in the asking field's vocabulary, the commercial index returns that field's own literature and continues to return it to the bottom of the list. Figure 1 and Table 6 give the share by rank band.

Figure 1. Share of each arm's sampled results that belong to the asking field's own literature, by rank band. The commercial index on the literal problem statement does not decline with depth.

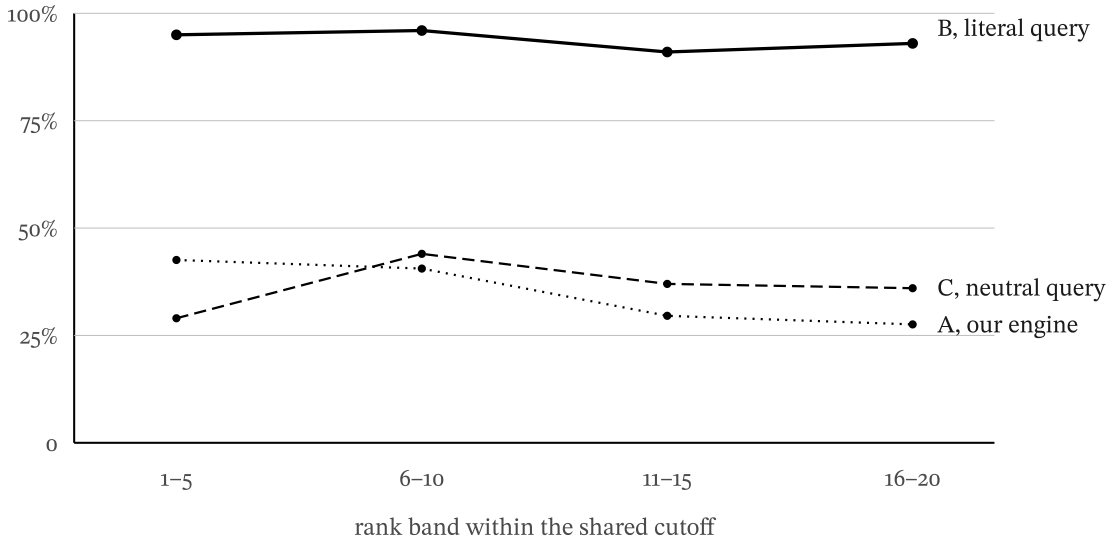

Table 6. Share of sampled results that belong to the asking field's own literature, by rank band, over the 30.3 percent sample inside the shared cutoff.

Ranks	A, our engine	B, literal query	C, neutral query
1 to 5	43%	**95%**	29%
6 to 10	41%	**96%**	44%
11 to 15	30%	**91%**	37%
16 to 20	28%	**93%**	36%

Arm B's share is flat across the list, so reading further down it does not reach other fields. In 37 of the 50 cases every result inside the cutoff is the asking field's own literature, and under the second adjudication instrument that figure is 46. This is the expected behavior of retrieval that ranks by similarity to the query, given a query written in the asking field's words.

The bound is not lifted by rewriting the query. Arm C sends our own domain-neutral restatement to the same index and its share drops to between 29 and 44 percent, close to ours. It reaches the field the historical case named in 2 of 50 cases, against 3 for the literal query and 21 for our engine. Leaving the asking field and arriving in the field that solved the problem are separate capabilities, and the query supplies only the first. §11.1 gives the evidence for why.

Table 7 reports the three case-level measures against that background. Arm B also returns the bridge paper for 31 of the 50 cases and 19 of those inside the cutoff, while reaching the field that transfer drew on in 3. The bridge paper is the write-up of the transfer and postdates the problem it is being returned for, so it is excluded from scoring on every arm: a system that returns it has retrieved the answer rather than the material the answer was built from. Given a problem posed before the connection existed, the commercial index finds the account of that connection far more often than it finds the field or the work the connection came from.

Table 7. Fifty cases, three arms, adjudicated blind over a 30.3 percent random sample of the results inside the shared cut-off.

Question	A, our engine	B, literal query	C, neutral query
1. Returned work from outside the asking field	**49 / 50**	13 / 50	39 / 50
returned *only* the asking field's own literature	1 / 50	37 / 50	11 / 50
individual results outside the asking field	192 / 300	18 / 300	190 / 300
2. Reached the field the historical case named	**21 / 50**	3 / 50	2 / 50
3. Of that out-of-field work, the share satisfying the criteria	**61 / 192**	5 / 18	34 / 190
as a rate	32%	28%	18%
cases with at least one paper satisfying the criteria	29 / 50	3 / 50	17 / 50

A fourth measure is reported separately in §11.7, because it rests on a different basis and does not belong in this table. It asks whether an arm returns the specific paper the transfer was built on, rather than the field that paper sits in. Our engine returns it in 10 of 43 cases and inside the cutoff in 7; both comparison arms return it in none. Those figures come from title matching against a catalogue rather than from the blind adjudication, over 43 cases rather than 50, and after the source papers had been added to our index. Table 10 and §A give the conditions.

On question 2 the difference between our engine and each comparison arm is 21 of 50 against 3 and 2, which is $p = 3.6 \times 10^{-5}$ and $p = 7 \times 10^{-6}$ by Fisher's exact test; the two comparison arms do not differ from each other. At the level of individual papers, 38 of our 192 out-of-field results sit in the field the case named against 3 of arm C's 190, $p = 1.9 \times 10^{-9}$. On question 3 the rates are 32, 28 and 18 percent, and the gap between our engine and the neutral query holds when papers are clustered by case, at 23 to 40 percent against 10 to 26 percent with design effects of 1.7 and 2.3. Question 1 sits at a ceiling for our engine, 90 to 100 percent, and separates the arms least.

The controls described in §8.3 run under the same checks and the same judge. The eight historical source papers score 8 of 8. The eleven deliberately wrong pairings score 0 of 11.

11.1 Rewriting the query leaves the field but does not aim

Arm C searches the same index as arm B. The one difference is the query: our engine's neutral restatement of the problem in place of the literal one.

On question 1 it behaves almost exactly like our engine: 190 of its 300 adjudicated results sit outside the asking field, against 192 for arm A and 18 for arm B. Rewriting the query in domain-neutral language is sufficient to leave the researcher's own literature.

On question 2 arm C reaches the field that historically solved the problem in 2 of 50 cases, which is indistinguishable from the literal query's 3 and one tenth of arm A's 21. On question 3, 18 percent of its out-of-field results satisfy the criteria, against 32 percent for arm A, over 17 cases against 29. Of arm C's 190 out-of-field papers, 3 sit in the field the case named and 187 in a third field.

The failure is visible in why the papers were rejected. Of arm C's 156 rejections, 78 are the effect failing to survive the change in conditions, against 42 of arm A's 131 over a pool of almost identical size. Arm C returns work that shares the shape of the problem and carries no mechanism that could cross. A further 33 fall on the precondition check, which is the same failure in its procedural form.

We attribute the gap to arm C neutralizing one side of the match only. Our engine applies the abstraction operator to the query and to every indexed document, so a neutral statement is matched against neutral statements. Arm C matches a neutral statement against literal text, and retrieves documents on vocabulary overlap with the restatement.

11.2 Where the out-of-field work landed

Leaving the asking field has two destinations. A paper lands either in the field the historical case named or in some third field that neither the case nor the researcher nominated. The two carry different pass rates, so a single out-of-field figure averages over them.

Table 8. The out-of-field papers split by destination, with the share of each that passes the checks.

	In the named field	Passing	In a third field	Passing
A, our engine	38	**23 (61%)**	154	38 (25%)
B, literal query	7	2	11	3
C, neutral query	3	2	187	32 (17%)

On third-field papers the arms sit between 17 and 27 percent, close enough that the difference between them is not carried there. They differ in how many papers reach the named field at all, 38 against 7 and 3, and in what happens to those papers: 61 percent of arm A's pass, against 25 percent of its own third-field work.

Arm C's 34 passing papers are more concentrated than arm A's, with the top three cases holding 38 percent against 23 percent. Four of them, which is 12 percent of its total, come from a single case where the asking field is ultrasound diagnosis of abdominal and pelvic masses and the papers returned are ultrasound of breast masses and of musculoskeletal cyst-like tumours. The clinical specialty differs while the imaging physics and the task of separating cystic from solid do not.

Three of arm C's passing papers do travel a real distance: Murray's law into clogging and permeability oscillations during mineral replacement, protein rigidity into linkage mechanisms in animal joints, and spin-glass dynamics into metastable states in neural circuits.

Index size is often offered as the explanation for a gap of this kind: a smaller corpus crowds a candidate less, so it is easier to surface. That account can be tested here, because landing in the field a case named can be compared against chance, and chance does not depend on how large the index is. A system that leaves the asking field and lands at random among N plausible fields hits the one field the case named with probability $1/N$, whatever the size of the corpus it drew from.

Table 9. Papers landing in the field the case named, against the number expected if out-of-field results were distributed at random over N fields. Arm A drew 192 out-of-field results and arm C drew 190, from indexes differing in size by roughly three orders of magnitude.

Fields, N	Expected by chance	A, our engine	C, neutral query
15	12.8	**38 (3.0×)**	3 (0.2×)
20	9.6	**38 (4.0×)**	3 (0.3×)
26	7.4	**38 (5.1×)**	3 (0.4×)
30	6.4	**38 (5.9×)**	3 (0.5×)

At $N = 20$ our engine is above chance at $p = 3.7 \times 10^{-13}$ by a binomial test, and at $N = 26$ at $p = 1.1 \times 10^{-16}$. Arm C is at or below chance at every value of N tested, $p = 1.00$ and $p = 0.98$ respectively. Two further figures bear on the same question. Arm B and arm C search the same index, so they face identical crowding, and their out-of-field shares are 6 and 63 percent; crowding therefore does not determine whether an arm leaves the asking field. And arm A and arm C return almost the same volume of out-of-field work, 192 against 190, from indexes three orders of magnitude apart; index size does not determine that volume either. What separates the arms is where the work lands, and that is measured against a baseline size does not move.

11.3 Adjacent pairs are not carrying the result

Seven of the fifty cases pair fields that sit close together, and a reader may reasonably ask whether they carry the difference between the arms. Scored over the 43 distant pairs alone, our engine leaves the asking field in 42 cases, reaches the field the case named in 18, and returns at least one paper satisfying the criteria in 27. The literal query gives 10, 1 and 2; the neutral query 33, 1 and 14. On question 2 that is 18 of 43 against 1 and 1, at $p = 9.9 \times 10^{-6}$ against each.

Removing the adjacent pairs moves our engine from 21 of 50 to 18 of 43 and both comparison arms from 3 and 2 to 1 and 1. The separation is wider on the distant pairs than on the full set, so the adjacent cases were reducing the measured difference rather than producing it.

11.4 What question 1 measures

Question 1 counts what left the asking field and question 3 counts how much of it satisfies the criteria. Of arm A's 192 out-of-field results, 61 pass, 107 fail, and 24 are disputed between the two assessors.

Both outcomes are visible in the same case. For a problem about protecting micromachined sensors from 60,000 g impacts, arm A returned vibration decay in a human skull after impact, impact resistance of fibre-reinforced composites, shock-proofing of small mobile displays, and dynamic dampers on an air mount: four fields, all of them about absorbing a hard hit. For a problem about telling a fluid-filled cyst from a solid tumour, it returned delamination cracks in textile composites and seismic lithofacies analysis alongside the industrial ultrasonics the case named.

The failures share one shape, which is a structural match carrying no mechanism. For a problem about the trade-off between water throughput and solute rejection in a desalination membrane, the arm returned thermoelectric materials twice. Thermoelectrics has a well-known trade-off between electrical and thermal conductivity, so the two problems have the same structure and no transferable mechanism. A coatings problem phrased around layer-by-layer deposition returned a deep-learning paper titled *A Layer-by-Layer Amplifier*. The abstraction matches on structure, and that produces the successes and these failures. The checks in §8.3 separate them.

11.5 Why question 2 understates

Each case names one field that historically solved the problem, and makes no claim that nothing else would. A battery-management case whose recorded answer is aerospace Kalman filtering returned state estimation for legged robots, which is the same structure in a third field and is scored as a miss. A magma crystallisation case returned the rate of change of boundary area in coarsening three-dimensional cellular structures, which is the same problem as crystal size distribution and answers the case at least as well as the field it named. Papers of that kind are counted by question 1 and by question 3, and only question 2 discards them.

11.6 Arm B, the literal query

Arm B returned nothing but the asking field's own literature in 37 of the 50 cases, and 18 of its 300 adjudicated results came from outside the asking field, against 192 for arm A.

Question 3 does not separate the two arms: 5 of 18 against 61 of 192 is $p = 0.80$ by Fisher's exact test. What separates them is how much out-of-field work each arm produced and how many cases it reached, 192 results against 18 and 29 cases against 3, both at $p < 10^{-7}$.

The papers it returns are on topic. For the crystallography case it returned CALYPSO and three other structure prediction codes, which is close to the ideal answer to the question *what has my field published about this*. Its 18 out-of-field results satisfy the criteria at 28 percent, against 32 percent for arm A.

11.7 The source paper itself

Two papers sit on opposite sides of each problem in time. The bridge paper is the write-up of the transfer, published after it was made; on the cardiology case that is Garfinkel and colleagues in 1992. The source paper is the work in field B the bridge paper drew on, published before the problem was posed; there, Ott, Grebogi and Yorke in 1990. A researcher facing the problem could have read the source paper. The bridge paper did not yet exist. Across the 47 identified cases the source paper predates its bridge paper in every one.

Questions 1 to 3 ask whether an arm reaches the field the source paper sits in. This question is narrower: whether it returns that paper, given only the problem as the asking field would have written it beforehand. Our engine does so in 10 of 43 cases, 7 of them inside the top twenty. Neither comparison arm returns it on any case at any depth.

Forty-seven of the fifty source papers were identified and each was verified against a live catalogue record: the title had to return from the catalogue, the year had to precede the bridge paper, the work had to sit in field B, and it had to be distinct from the bridge paper. Where the bridge paper's own reference list was available it was read, and most identifications were confirmed by finding the candidate cited there. Kantorovich for optimal transport into cell-fate trajectories, Waite 1981 for mussel adhesion into coatings, and the capillary-wave measurement for P granules into phase separation were each confirmed that way.

Forty-three carried an abstract in at least one public source and could be ingested. Four did not and were dropped.

Before this test the index held one of the fifty source papers, which is also the condition under which the 21 of 50 in Table 7 was measured. An engine cannot return a paper it does not have, so the narrower question was unanswerable as the index stood. The forty-three were ingested through the normal path, which generated their abstractions and embeddings exactly as it does for every other paper, and the fifty problem statements were then re-run unchanged.

The comparison arm was given no equivalent step, so we probed its index directly instead. Each of the 43 source papers was queried by its exact title, one call per paper. All 43 are present: 37 are returned at rank 1 and the median title match is exact. The comparison arm's 0 of 43 in Table 10 is therefore a ranking result and not an absence. Every one of those papers is in the index it searched, retrievable by a query that names it, and no problem statement reached one.

The same run shows the index holds the bridge papers as well: it returns the bridge paper for 31 of the 50 cases and 19 of those inside the cut-off, against 8 and 2 for arm C and 1 and 1 for our engine. Those are anachronisms rather than answers, which is why the bridge paper is excluded from scoring on every arm.

Table 10. Does the arm return the paper the transfer borrowed from. All three arms are scored over the same 43 cases, the ones whose source paper carried an abstract and could be ingested; arms B and C also return none of the other four. Our pool is 120 deep and the comparison arm was asked for 100, which does not affect the count, since our ten hits rank between 1 and 72. Matching is by title against the catalogue record, and every candidate match was read by hand. All 43 papers are present in both indexes: they were added to ours for this measurement, and a direct title probe returns all 43 from the comparison index, 37 of them at rank 1. Each arm was then given the problem statement alone. The two indexes differ in size by roughly three orders of magnitude, which §A.6 records.

	In the pool	Inside the cut-off of 20
A, our engine	**10 / 43**	**7 / 43**
B, literal query	0 / 43	0 / 43
C, neutral query	0 / 43	0 / 43

Two of the ten, with the input each arm was given. The cardiology case supplies a description of an irregular heartbeat in cardiology's vocabulary as of the period before 1992, and the engine returns Ott, Grebogi and Yorke's 1990 chaos-control paper at rank 7. The MEMS case supplies a packaging problem for micromachined sensors that have to survive about 60,000 g of shock. That statement names no animal, no skull and no bone; the engine returns the mechanical analysis of how woodpeckers avoid brain injury at rank 1. Both statements are in the case file and both pass leak rule 1.

The two apparent hits on arm B were read and neither is the source paper. One is the bridge paper itself, the other a later applications paper sharing most of a short title. The count was checked at four matching thresholds and does not move until the threshold is loose enough to admit those two.

Table 11. Arm A's recovery by transfer type, and the rank at which the source paper came back.

Type	Cases	Returned	Inside 20	Ranks
MECHANISM	18	6	4	1, 4, 4, 9, 48, 72
INSTRUMENT	4	2	2	3, 20
METHOD	16	2	1	7, 26
PRACTICE	5	0	0	

Eight of the ten are mechanism or instrument transfers, 8 of 22 such cases, against 2 of 21 method and practice cases. On these counts that difference is $p = 0.069$ by Fisher's exact test, so it is a direction and not an established effect. The same direction appears in the usefulness checks, where the precondition test rejects 15 of arm A's 192 out-of-field papers, 1 of arm B's 18 and 33 of arm C's 190, and rejects none of the eight controls.

The precondition test is what produced that ordering in the checks. Before it was added, method transfers passed at 50 percent against 33 for mechanism transfers, since a check asking whether an effect survives a change in conditions has no purchase on a procedure. Table 11 was measured separately.

12 Conclusions

Cross-field retrieval could not be scored because it had no test collection. Fifty documented transfers supply one, and the judgments in it were made by other researchers, published before this work, and are checkable from Appendix B without access to any index.

Measured against it, retrieval that ranks by similarity to the query does not perform the task at any depth. The commercial index returns the asking field’s own literature in 91 to 96 percent of results across every rank band and returns nothing else in 37 of the 50 cases. That is the expected behavior of the method given a query in the asking field’s words, and it bounds what the task can get from it.

Rewriting the query does not lift the bound. Arm C, on that same index, leaves the asking field about as often as our engine does, 190 adjudicated papers against 192, and reaches the field that solved the problem in 2 of 50 cases against our 21. The query moves a search out of one literature; it does not direct it into another.

Abstracting the query and the corpus reaches the field the historical case named in 21 of 50 cases, and 61 of its 192 out-of-field results satisfy criteria that score eight historical source papers 8 of 8 and eleven mismatched pairings 0 of 11. It misses more cases than it reaches. What the measurement establishes is that the task is not out of reach of retrieval, and that the part of the method which supplies the aim sits on the corpus side rather than in the query.

Whether retrieval of this kind changes the rate at which such transfers occur is outside what fifty historical cases can establish. What they establish is that the source work is reachable from the problem statement: rank 7 for the 1990 chaos-control paper given the cardiology problem as it stood before 1992, rank 1 for the woodpecker biomechanics given a packaging problem for micromachined sensors.

Three experiments follow from this one. Running the comparison arm against an index the size of ours would separate the ranking from the crowding, which the present design leaves together (§A.5, §A.6). Having the problem statements written by domain experts who do not know the transfer would remove the last dependence on an author who did (§A.1). And a human panel over a sample of the 900 adjudications would put the relevance estimates on something other than a model (§A.2). The protocol in §13 is specified so that a group with no access to our system can run it against any retrieval service, including ours.

13 Reproduction

The protocol is specified above and is designed to be run by a group with no access to our system. What a third party needs is the case set, the ordering of the steps, and the property each step has to hold.

1. **Cases.** Fifty transfers, each recorded with an asking field, a source field, a problem statement in the asking field's vocabulary as of the period before the transfer, five terms native to the source field, and the

bridge paper. Each problem statement passes the leak audit in §3 before it is used.

2. **Retrieval.** Each arm receives the problem statement and nothing else, is asked for the same number of results and is scored to the same cut-off. The bridge paper is excluded on every arm. A case that returns nothing aborts the run; an empty record is never written.
3. **Corpus probe.** Before each case is searched, the index is checked directly for that case's target terms by exact substring match, so a coverage miss is never scored as a ranking loss.
4. **Sampling.** Results inside the cut-off are ordered by a hash of a stable identifier and the first *n* per case per arm are taken, so the selection is reproducible and no scorer takes any part in it.
5. **Blinding.** The adjudication set carries identical fields for every item and no arm label. Abstracts come from public sources through one pipeline for every arm, and coverage across arms is asserted to within 15 points before adjudication starts.
6. **Adjudication.** Two instruments, differing in how the question is put, decide which community each paper belongs to over the whole set. The agreement between them is reported.
7. **Checks.** Every out-of-field paper goes through the checks in §7.3, with the positive and the negative control run alongside on the same judge.
8. **Scoring.** The scorer refuses to report a figure over an incomplete verdict file.

The case set and the harness are available on request for replication.

13.1 Reproducing the source-paper measurement

1. **Identify.** For each case, name the earlier publication in field B that the bridge paper drew on. Verify each against a public catalogue: it must return by title, predate the bridge paper, sit in field B, and differ from the bridge paper. Read the bridge paper’s reference list where the catalogue exposes it and prefer a candidate found there.

2. **Ingest.** Add the identified papers to the index through the ordinary path, so that they receive the same processing as every other paper. Record which artifacts were added, so the result can be computed with and without them.

3. **Re-run.** Submit the same fifty problem statements, unchanged, under the deployed engine settings. Record the full ranked list, not just whether the paper appeared.

4. **Probe the comparison index.** Query each source paper by its exact title against every index under test, one call per paper, and record whether it comes back and at what rank. An arm that cannot return a paper for its own title cannot return it for anything, and a zero on the ranked task means something different in that case.

5. **Match.** Resolve each source paper to an artifact identifier and search the ranked list by identifier, not by title text. Where title matching is used for the other arms, read every candidate match; short titles produce false positives against both the bridge paper and later applications.

6. **Report.** Give the position in the pool and the position inside the cut-off separately. The first is what a reading layer could recover, the second is what a person scanning results would see.

Appendix A. Limitations

No benchmark is complete. We list below the properties of this one that bound what its figures support, and the experiments that would remove each bound.

1. **Fifty cases, in English, written by hand.** The set covers biology into engineering, physics into biology, mathematics into molecular genetics, machine learning into the physical sciences, and operational practice between industries. It is not a sample of transfers in general, and famous transfers such as the gecko and mussel cases carry decades of downstream literature that makes them easier for every arm. Enlarging the set, and having the problem statements written by people who do not know the answer, are the next things it needs.

2. **Adjudication is model-based.** The four checks in Table 4 establish that an instrument is stable, order-invariant, blind and correct on a fixed set assembled in advance; none establishes that it is correct on the items it is about to score. The design answers to that with two differently-posed instruments over all 900 items, published disagreement, a positive and a negative control, and no figure produced by the model family the engine under test runs on. A human panel over a sample would answer it better.

3. **The checks read abstracts.** A paper whose conditions live in its methods section can fail on terseness. 177 of the 900 items carry no abstract from any source and are judged from the title. Full text would move question 3 in a direction this run cannot predict.

4. **One draw per case.** The abstraction operator samples, so two runs of one case retrieve overlapping but different pools: 64 of 120 papers and 9 of the top 20 between consecutive draws. Every figure here is one draw. On a fourteen-case subset, unioning three draws raised source-paper recovery inside the cut-off from 2 to 5, so repeated draws are more likely to raise these numbers than to lower them.

5. **The two indexes are not the same size.** The comparison arm searches roughly three orders of magnitude more papers. That supplies it with far more of the asking field's own work for an asking-field query to rank highly, so part of its 37 of 50 returning only that literature may be a property of index size acting through similarity ranking. Table 9 bounds what that can explain: landing in the named field is measured against a chance baseline that does not depend on index size, and arm C is at chance on the larger index while arm A is three to six times above it. Index size is therefore not available as an account of the difference on question 2. Running the comparison arm against an index the size of ours remains the experiment that would settle question 1 and the ranking result in Table 10.

6. **Table 10 is not size-controlled.** Both indexes contain all 43 source papers, so the comparison arm’s zero is a ranking result rather than a coverage one. The indexes still differ in size by roughly three orders of magnitude, so the 43 face far fewer competing documents in ours. Separating the ranking from the crowding requires the same experiment as item 5.

7. **Sampling.** Question 3 and the destination split rest on 30.3 percent of the results inside the cut-off. The case-level counts are lower bounds, since sampling can hide a case’s only out-of-field result and cannot

invent one; the rates are two-sided estimates, reported above with intervals. Six sampled results per case per arm attenuate the case-level counts more for an arm with a low out-of-field density than for a high one, which is a further reason question 1 is the weakest of the three.

Appendix B. The case set

All fifty cases, with the publication that records each transfer. The asking field and the source field are reproduced from the case file without abbreviation. A reader can check any case against its DOI without access to the harness or to either index.

#	Asking field → source field	Publication recording the transfer
1	Archaeology and palaeoanthropology (site geochronology) → Physical organic chemistry and stereochemistry (racemization kinetics)	Jeffrey L. Bada, Reiner Protsch, "Racemization Reaction of Aspartic Acid and Its Use in Dating Fossil Bones", Proceedings of the National Academy of Sciences USA, 1973, vol. 70, no. 5, pp. 1331-1334. 10.1073/pnas.70.5.1331
2	Combinatorial optimisation and computer-aided physical design of digital computers (partitioning, component placement, wiring) → Statistical mechanics of condensed matter, and the metallurgy of annealing	S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, "Optimization by Simulated Annealing", Science 220(4598):671-680, 13 May 1983. 10.1126/science.220.4598.671
3	Petroleum production chemistry and subsea pipeline flow assurance → Cold-adaptation physiology of fish and insects (antifreeze proteins)	Shixin Li, Rujie Lv, Zengshuai Yan, Fang Huang, Xianren Zhang, Guangjin Chen, Tongtao Yue, "Design of Alanine-Rich Short Peptides as a Green Alternative of Gas Hydrate Inhibitors: Dual Methyl Group Docking for Efficient Adsorption on the Surface of Gas Hydrates", ACS Sustainable Chemistry & Engineering, 2020, vol. 8, no. 10, pp. 4256-4266. 10.1021/acssuschemeng.9b07701
4	Membrane separation engineering (water treatment and desalination) → Membrane physiology and structural biology (aquaporin water channels)	Manish Kumar, Mariusz Grzelakowski, Julie L. Zilles, Mark M. Clark, Wolfgang Meier, "Highly permeable polymeric membranes based on the incorporation of the functional water channel protein Aquaporin Z", Proceedings of the National Academy of Sciences USA, 2007, vol. 104, no. 52, pp. 20719-20724. 10.1073/pnas.0708762104
5	Asthma and airway epithelial biology (respiratory medicine) → Soft-matter and granular physics: the jamming/unjamming transition of close-packed inert matter	Jin-Ah Park, Jae Hun Kim, Dapeng Bi, Jennifer A. Mitchel, et al. (incl. M. Lisa Manning, Jeffrey M. Drazen, Jeffrey J. Fredberg), "Unjamming and cell shape in the asthmatic airway epithelium," Nature Materials 14(10):1040-1048, 2015. 10.1038/nmat4357
6	Anaesthesiology (management of rare intraoperative crises) → Commercial and military aviation flight-crew training	Gaba DM, Howard SK, Fish KJ, Smith BE, Sowb YA. "Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience." Simulation & Gaming 2001;32(2):175-193. 10.1177/104687810103200206
7	Cardiology / cardiac electrophysiology → Nonlinear dynamics and chaos control (the Ott-Grebogi-Yorke method of stabilising unstable periodic orbits)	Alan Garfinkel, Mark L. Spano, William L. Ditto & James N. Weiss, "Controlling Cardiac Chaos," Science 257(5074):1230-1235, 1992. 10.1126/science.1519060

#	Asking field → source field	Publication recording the transfer
8	Animal digestive physiology and feeding ecology → Chemical reaction engineering (ideal reactor theory)	Deborah L. Penry, Peter A. Jumars, "Modeling Animal Guts as Chemical Reactors", The American Naturalist, 1987, vol. 129, no. 1, pp. 69-96. 10.1086/284623
9	RF integrated circuit design and wideband spectrum analysis → Auditory physiology (mammalian hearing)	Soumyajit Mandal, Serhii M. Zhak, Rahul Sarpeshkar, "A Bio-Inspired Active Radio-Frequency Silicon Cochlea", IEEE Journal of Solid-State Circuits 44(6):1814-1828 (2009) 10.1109/JSSC.2009.2020465
10	Structural / civil engineering - damage and stress monitoring of loaded reinforced and post-tensioned concrete members → Seismology - coda (multiply scattered) wave interferometry for detecting sub-percent velocity changes in the Earth	Ernst Niederleithinger, Xin Wang, Martin Herbrand, Matthias Muller (2018). "Processing Ultrasonic Data by Coda Wave Interferometry to Monitor Load Tests of Concrete Beams." Sensors 18(6), 1971. 10.3390/s18061971
11	Magnetic resonance imaging - clinical MR acquisition and scan-time reduction → Compressed sensing - applied harmonic analysis and information theory (mathematics)	Michael Lustig, David Donoho, John M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging", Magnetic Resonance in Medicine 58(6):1182-1195, 2007. 10.1002/mrm.21391
12	Neuroscience (spontaneous cortical network activity) → Statistical physics of self-organized criticality - sandpile avalanche models, earthquake statistics, critical branching processes	John M. Beggs & Dietmar Plenz, "Neuronal Avalanches in Neocortical Circuits," The Journal of Neuroscience 23(35):11167-11177, 2003. 10.1523/JNEUROSCI.23-35-11167.2003
13	Respiratory-infection transmission and public-health precautions → Multiphase fluid mechanics (turbulent buoyant momentum puffs, entrainment, sedimentation from turbulent flows)	Lydia Bourouiba, Eline Dehandschoewercker & John W. M. Bush, "Violent expiratory events: on coughing and sneezing," Journal of Fluid Mechanics 745:537-563, 2014. 10.1017/jfm.2014.88
14	Seismology - detecting and cataloguing induced earthquakes → Artificial intelligence / computer vision - deep convolutional neural networks	Thibaut Perol, Michael Gharbi, Marine Denolle, "Convolutional neural network for earthquake detection and location", Science Advances 4(2):e1700578, 2018. 10.1126/sciadv.1700578
15	Microbial ecology (bacterial attachment to surfaces) → Colloid	Mark C. M. van Loosdrecht, J. Lyklema, Willem Norde, Alexander J. B. Zehnder, "Bacterial adhesion: A physicochemical approach", Microbial Ecology, 1989,

#	Asking field → source field	Publication recording the transfer
	and interface chemistry (DLVO theory of colloidal stability)	vol. 17, no. 1, pp. 1-15. 10.1007/BF02025589
16	Molecular genetics and enzymology of site-specific recombination → Knot theory / low-dimensional topology (mathematics)	Steven A. Wasserman, Jan M. Dungan & Nicholas R. Cozzarelli, "Discovery of a Predicted DNA Knot Substantiates a Model for Site-Specific Recombination," Science 229(4709):171-174, 1985. 10.1126/science.2990045
17	Volcanology - simulating explosive eruption columns, pyroclastic flows and surges → Fossil-energy / chemical process engineering - gas-solid multiphase reactor simulation (MFIX, built at DOE's fossil-energy laboratory for fluidized-bed coal reactors)	Sebastien Dartevelle, William I. Rose, John Stix, Karim Kelfoun, James W. Vallance (2004). "Numerical modeling of geophysical granular flows: 2. Computer simulations of plinian clouds and pyroclastic flows and surges." Geochemistry, Geophysics, Geosystems, vol. 5. 10.1029/2003GC000637
18	Petroleum reservoir engineering - continuous updating of a near-well reservoir model and its production forecast → Physical oceanography - sequential data assimilation into large non-linear ocean circulation models (Evensen's ensemble Kalman filter)	Geir Naevdal, Trond Mannseth, Erlend H. Vefring (2002). "Near-Well Reservoir Monitoring Through Ensemble Kalman Filter." SPE/DOE Improved Oil Recovery Symposium, Tulsa, Oklahoma, April 2002, paper SPE-75235-MS. 10.2118/75235–MS
19	Software engineering practice (industrial technology adoption and process improvement) → Clinical medicine (evidence-based medicine and its appraisal machinery)	Dyba T, Kitchenham BA, Jorgensen M. "Evidence-Based Software Engineering for Practitioners." IEEE Software 2005;22(1):58-65. 10.1109/MS.2005.6
20	Crystallography, solid-state chemistry and high-pressure mineral physics → Evolutionary computation / genetic algorithms (computer science)	Artem R. Oganov, Colin W. Glass, "Crystal structure prediction using ab initio evolutionary techniques: Principles and applications", The Journal of Chemical Physics 124(24):244704, 2006. 10.1063/1.2210932
21	Paediatric cardiac surgery and paediatric intensive care (clinical handover) → Motorsport race-team pit-crew operations (Formula 1), with commercial aviation flight-crew procedure design	Catchpole KR, de Leval MR, McEwan A, Pigott N, Elliott MJ, McQuillan A, MacDonald C, Goldman AJ. "Patient handover from surgery to intensive care: using Formula 1 pit-stop and aviation models to improve safety and quality." Pediatric Anesthesia 2007;17(5):470-478. 10.1111/j.1460–9592.2006.02239.x
22	Organismal biology / functional morphology and comparative	Kellar Autumn, Metin Sitti, Yiching A. Liang, Anne M. Peattie, Wendy R. Hansen, Simon Sponberg, Thomas W. Kenny, Ronald Fearing, Jacob N.

#	Asking field → source field	Publication recording the transfer
	biomechanics of lizards → Surface physics and intermolecular forces (van der Waals dispersion forces, contact mechanics, contact splitting)	Israelachvili & Robert J. Full, "Evidence for van der Waals adhesion in gecko setae," Proceedings of the National Academy of Sciences USA 99(19):12252-12256, 2002. 10.1073/pnas.192252799
23	Microfabrication and adhesive materials engineering → Reptile functional morphology / vertebrate locomotion biology	A. K. Geim, S. V. Dubonos, I. V. Grigorieva, K. S. Novoselov, A. A. Zhukov, S. Yu. Shapoval, "Microfabricated adhesive mimicking gecko foot-hair", Nature Materials 2(7):461-463 (2003) 10.1038/nmat917
24	Igneous petrology and volcanology - reading crystallization kinetics and magma solidification history out of rock textures → Chemical engineering - industrial crystallizer design and the population-balance treatment of crystal populations	Bruce D. Marsh (2007). "Crystallization of Silicate Magmas Deciphered Using Crystal Size Distributions." Journal of the American Ceramic Society, vol. 90. 10.1111/j.1551-2916.2006.01473.x
25	Obstetrics and gynaecology (diagnosis of abdominal and pelvic masses) → Industrial ultrasonic flaw detection and non-destructive testing of metals and welds	Donald I, MacVicar J, Brown TG. "Investigation of abdominal masses by pulsed ultrasound." The Lancet 1958;1(7032):1188-1195. 10.1016/S0140-6736(58)91905-6
26	Electrochemical energy storage - hybrid-electric-vehicle battery pack engineering and battery management systems → Stochastic estimation and control theory - Kalman filtering, matured in aerospace navigation and guidance	Gregory L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 1: Background", Journal of Power Sources 134:252-261, 2004. 10.1016/j.jpowsour.2004.02.031
27	MEMS sensor engineering and underwater vehicle navigation → Fish sensory neurobiology	Yingchen Yang, Jack Chen, Jonathan Engel, Saunvit Pandya, Nannan Chen, Craig Tucker, Sheryl Coombs, Douglas L. Jones, Chang Liu, "Distant touch hydrodynamic imaging with an artificial lateral line", PNAS 103(50):18891-18895 (2006) 10.1073/pnas.0609274103
28	Biogerontology and ophthalmology (age- and diabetes-related damage to long-lived proteins) → Food chemistry (nonenzymatic browning of stored and cooked foods)	Vincent M. Monnier, Anthony Cerami, "Nonenzymatic Browning in Vivo: Possible Process for Aging of Long-Lived Proteins", Science, 1981, vol. 211, no. 4481, pp. 491-493. 10.1126/science.6779377
29	Oncology and cancer nanomedicine → Metallurgy of amorphous alloys (metallic glasses)	Chen Zhang, Wenbo Bu, Dalong Ni, Shenjian Zhang, Qing Li, Zhenwei Yao, Jiawen Zhang, Heliang Yao, Zheng Wang, Jianlin Shi, "Synthesis of Iron Nanometallic Glasses and Their Application in Cancer Therapy by a Localized

#	Asking field → source field	Publication recording the transfer
		Fenton Reaction", Angewandte Chemie International Edition, 2016, vol. 55, no. 6, pp. 2101-2106. 10.1002/anie.201510031
30	Optical engineering (lens antireflection treatment) → Entomology and insect vision biology	P. B. Clapham, M. C. Hutley, "Reduction of Lens Reflexion by the 'Moth Eye' Principle", Nature 244(5414):281-282 (1973) 10.1038/244281a0
31	Porous materials, catalysis and electrochemical energy engineering → Physiology and plant biology (biological transport networks)	Xianfeng Zheng, Guofang Shen, Chao Wang, Yu Li, Darren Dunphy, Tawfique Hasan, C. Jeffrey Brinker, Bao-Lian Su, "Bio-inspired Murray materials for mass transfer and activity", Nature Communications 8:14921 (2017) 10.1038/ncomms14921
32	Surface and materials chemistry (thin-film coatings and surface functionalization) → Marine invertebrate biology (byssal adhesion in mussels)	Haeshin Lee, Shara M. Dellatore, William M. Miller, Phillip B. Messersmith, "Mussel-Inspired Surface Chemistry for Multifunctional Coatings", Science, 2007, vol. 318, no. 5849, pp. 426-430. 10.1126/science.1147241
33	Clinical bone assessment and osteoporotic fracture risk → Non-destructive testing and evaluation of engineering waveguide structures (plates, pipes, rails)	Moilanen P. "Ultrasonic guided waves in bone." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 2008;55(6):1277-1286. 10.1109/TUFFC.2008.790
34	Computational chemistry and condensed-matter physics - first-principles molecular dynamics and interatomic potentials → Machine learning - artificial neural networks as general function approximators (computer science)	Jorg Behler, Michele Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces", Physical Review Letters 98(14):146401, 2007. 10.1103/PhysRevLett.98.146401
35	Machine learning - unsupervised generative modelling of high-dimensional data → Non-equilibrium statistical physics and thermodynamics	Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, Surya Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics", Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:2256-2265, 2015. arXiv:1503.03585
36	Developmental and stem-cell biology (iPSC reprogramming) → Optimal transport theory in mathematics (Monge-Kantorovich problem, Wasserstein couplings, entropic regularisation)	Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, et al. (incl. Philippe Rigollet, Aviv Regev, Eric S. Lander), "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming," Cell 176(4):928-943.e22, 2019. 10.1016/j.cell.2019.01.006
37	Cell and developmental biology (germ-cell specification in C.	Clifford P. Brangwynne, Christian R. Eckmann, David S. Courson, Agata Rybarska, Carsten Hoege, Jobin Gharakhani, Frank Julicher & Anthony A.

#	Asking field → source field	Publication recording the transfer
	elegans) → Soft-matter physics and physical chemistry of liquids and phase separation (surface tension, viscosity, droplet coalescence, condensation/dissolution)	Hyman, "Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation," Science 324(5935):1729-1732, 2009. 10.1126/science.1172046
38	Materials science and condensed-matter physics of glasses - silica glass, Lennard-Jones glass, Cu-Zr metallic glass → Computational algebraic topology - persistent homology and persistence diagrams	Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, Yasumasa Nishiura, "Hierarchical structures of amorphous solids characterized by persistent homology", PNAS 113(26):7035-7040, 2016. 10.1073/pnas.1520877113
39	Molecular biology and genetic diagnostics (DNA sequence analysis) → Semiconductor microfabrication (photolithography)	Ann Caviani Pease, Dennis Solas, Elizabeth J. Sullivan, Maureen T. Cronin, Christopher P. Holmes, Stephen P. A. Fodor, "Light-generated oligonucleotide arrays for rapid DNA sequence analysis", Proceedings of the National Academy of Sciences USA, 1994, vol. 91, no. 11, pp. 5022-5026. 10.1073/pnas.91.11.5022
40	Surface engineering / liquid-repellent and anti-fouling coatings → Botany (carnivorous plant surface morphology)	Tak-Sing Wong, Sung Hoon Kang, Sindy K. Y. Tang, Elizabeth J. Smythe, Benjamin D. Hatton, Alison Grinthal, Joanna Aizenberg, "Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity", Nature 477(7365):443-447 (2011) 10.1038/nature10447
41	Structural biology / computational biochemistry of proteins → Condensed-matter physics of covalent network glasses and combinatorial rigidity theory	Donald J. Jacobs, A. J. Rader, Leslie A. Kuhn & M. F. Thorpe, "Protein flexibility predictions using graph theory," Proteins: Structure, Function, and Bioinformatics 44(2):150-165, 2001. 10.1002/prot.1081
42	Nuclear fusion - tokamak plasma physics and magnetic-confinement control engineering → Deep reinforcement learning (machine learning / artificial intelligence)	Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, et al., "Magnetic control of tokamak plasmas through deep reinforcement learning", Nature 602(7897):414-419, 2022. 10.1038/s41586-021-04301-9
43	Aerospace non-destructive evaluation - large-area inspection of aircraft wing and fuselage skin → Seismology / exploration geophysics - reconstruction of subsurface structure from many crossing wave paths	Willi Schwarz, Michael Read, Matthew J. Kremer, Mark K. Hinders, Barry T. Smith (1999). "Lamb wave tomographic imaging system for aircraft structural health assessment." Proceedings of SPIE (Nondestructive Evaluation of Aging Aircraft, Airports and Aerospace Hardware III). 10.1117/12.339898
44	Health care quality and clinical practice measurement → Industrial quality management in electronics and diversified manufacturing	Chassin MR. "Is health care ready for Six Sigma quality?" The Milbank Quarterly 1998;76(4):565-591. 10.1111/1468-0009.00106

#	Asking field → source field	Publication recording the transfer
45	Protein biochemistry / molecular biophysics of protein folding → Condensed-matter statistical physics of disordered magnetic systems (spin-glass theory)	Joseph D. Bryngelson & Peter G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," Proceedings of the National Academy of Sciences USA 84(21):7524-7528, 1987. 10.1073/pnas.84.21.7524
46	Cardiac surgery and perfusion (intraoperative team communication) → Aviation flight-deck human factors and aerospace workload measurement	Wadhera RK, Parker SH, Burkhart HM, Greason KL, Neal JR, Levenick KM, Wiegmann DA, Sundt TM 3rd. "Is the 'sterile cockpit' concept applicable to cardiovascular surgery critical intervals or critical events? The impact of protocol-driven communication during cardiopulmonary bypass." Journal of Thoracic and Cardiovascular Surgery 2010;139(2):312-319. 10.1016/j.jtcvs.2009.10.048
47	Fibre-reinforced composite structures engineering (aerospace, automotive, armour) → Marine crustacean biology (stomatopod functional morphology)	L. K. Grunenfelder, N. Suksangpanya, C. Salinas, G. Milliron, N. Yaraghi, S. Herrera, K. Evans-Lutterodt, S. R. Nutt, P. Zavattieri, D. Kisailus, "Bio-inspired impact-resistant composites", Acta Biomaterialia 10(9):3997-4008 (2014) 10.1016/j.actbio.2014.03.022
48	Hospital medicine and hospital care delivery → Automotive manufacturing (Toyota Production System / lean production)	Kim CS, Spahlinger DA, Kin JM, Billi JE. "Lean health care: what can hospitals learn from a world-class automaker?" Journal of Hospital Medicine 2006;1(3):191-199. 10.1002/jhm.68
49	Polymer coatings and structural materials engineering → Vascular physiology / tissue repair biology	Kathleen S. Toohey, Nancy R. Sottos, Jennifer A. Lewis, Jeffrey S. Moore, Scott R. White, "Self-healing materials with microvascular networks", Nature Materials 6(8):581-585 (2007) 10.1038/nmat1934
50	MEMS packaging / high-g shock protection engineering → Ornithology and avian cranial anatomy	Sang-Hee Yoon, Sungmin Park, "A mechanical analysis of woodpecker drumming and its application to shock-absorbing systems", Bioinspiration & Biomimetics 6(1):016003 (2011) 10.1088/1748-3182/6/1/016003