APPENDIX A
PROMPTS

When constructing contrastive learning training data, we use GPT's classification to obtain L_{pos} and L_{neg}. For each entity in the top-T of L_0, determine whether its attributes are consistent with S^{pos} (S^{neg}). If it is consistent, GPT should output 1; otherwise, GPT should output 0. Next, these entities that GPT deems consistent with the attributes of S^{pos} (S^{neg}) will be merged with S^{pos} (S^{neg}) to form L_{pos} (L_{neg}). The prompt to classify entity is shown in Table XI.

In GenExpan, each round uses three example entities to construct $Prompt_g$, which is used to generate entities that are semantically similar to them. $Prompt_g$ is as shown in Table XII.

In the process of enhancing GenExpan with Chain of Thought prompting, we employ LLM to generate fine-grained class names. The prompt, $Prompt_c$, is shown in Table XIII.

APPENDIX B
IMPLEMENTATION DETAILS OF RETEXPAN

In RetExpan, sentences are tokenized using the WordPiece tokenizer and then fed into a 12-layer Transformer initialized with $BERT_{BASE}$ weights. To optimize training efficiency and preserve semantic knowledge learned by BERT, we freeze the first 11 layers of the encode, thus only the last layer is fine-tuned. When tokenizing, to ensure that mentions of entities in long sentences are not truncated, we have implemented entity focus, guaranteeing the presence of entities within the sentence. During expansion, we use one-shot expansion to directly obtain the preliminary expansion results L_0. In RetExpan with contrastive learning strategy, each epoch during training alternates between computing and optimizing entity prediction loss and contrastive learning loss.

To ensure that the encoder acquires knowledge from the corpus, we trained it for 20 epochs on 8 RTX 3090 GPUs. The hyperparameters for training learning rate, batch size, weight decay, label smoothing factor η were set to 4e-5, 128, 1e-2, and 0.075 respectively.

APPENDIX C
IMPLEMENTATION DETAILS OF GENEXPAN

In GenExpan, we use LLaMA-7b as the base model. Initially, we train for 1 epoch on the corpus using 6 A100 GPUs. The training hyperparameters learning rate, batch size, gradient accumulation steps, weight decay, gradient clipping are set to 1e-5, 4, 8, 1e-4, and 1.0 respectively.

During entity generation in GenExpan, we utilize prefix-constrained beam search with a beam size of 40 to generate 40 entities at a round. For entity selection, we compute the positive similarity score for each entity and select those whose scores are in the top 0.7 as the results for the current round. When no new entity is generated for 20 consecutive rounds, the generation process will end and move to re-ranking.

TABLE IX
TYPES OF ULTRA-FINE-GRAINED SEMANTIC CLASSES. CLS.: SEMANTIC CLASS

$\lvert\mathcal{A}^{pos}\rvert$	$\lvert\mathcal{A}^{neg}\rvert$	#Ultra-fine-grained CLS.
1	1	238
1	2	5
2	1	9
2	2	7
3	3	2

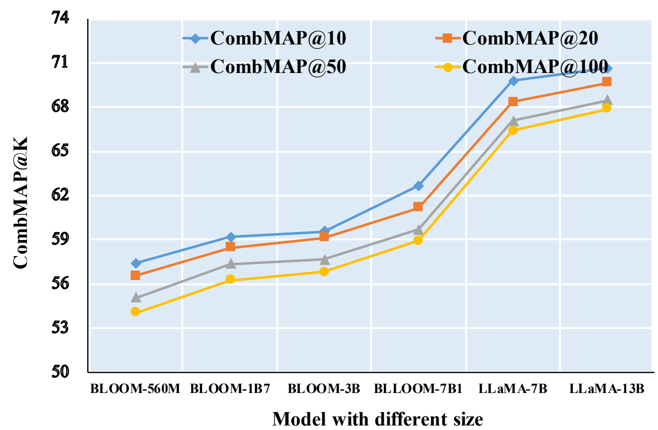

Fig. 8. Comparison experiments on different LLM families and sizes.

APPENDIX D
DETAILS OF ULTRAWIKI

UltraWiki contains 10 fine-grained semantic classes, which fall under 5 coarse-grained semantic categories: Organization, Location, Product, Person, and Miscellaneous. The number of entities in each fine-grained semantic class ranges from 45 to 952. Each fine-grained semantic class has 2 to 3 attributes, and depending on the combination of attributes, several ultra-fine-grained semantic classes can be derived from a single fine-grained class. Detailed data is displayed in Table X.

The number of combinations from positive and negative attributes varies significantly, leading to substantial differences in the number of ultra-fine-grained semantic classes. The count of UltraWiki's positive and negative attributes and their corresponding ultra-fine-grained classes are presented in Table IX.

APPENDIX E
ANALYSIS OF MODEL SIZE

We utilize various families and sizes of LLMs as backbone model of GenExpan, including 560M, 1B7, 3B, 7B1 for BLOOM, as well as 7B and 13B for LLaMA. The experiment results depicted in Figure 8 align well with the expectation. Both the BLOOM and LLaMA families satisfy the scaling law: larger models tend to be more effective in Ultra-ESE, yielding better results. Moreover, LLaMA-7B outperforms BLOOM-7B1 at the same scale.

TABLE X

FINE-GRAINED SEMANTIC CLASSES DETAIL. CLS.: SEMANTIC CLASS

Coarse CLS.	Fine-grained CLS.	#Entities	#Ultra-fine-grained CLS.	Attributes
Organization	Canada universities	99	10	<Loc-Province>, <Type>
Location	China cities	675	50	<Province>, <Prefecture>
	Countries	190	68	<Continent>, <Driving-Side>, <Per-Capita-Income>
	US airports	370	74	<Role>, <Loc-State>
	US national monuments	112	12	<Loc-State>, <Agency>
Product	Mobile phone brands	159	7	<Loc-Continent>, <Status>
	Percussion instruments	128	10	<Type>, <Source-Continent>
Person	Nobel laureates	952	11	<Prize>, <Gender>
	US presidents	45	5	<Party>, <Birth-State>
Miscellaneous	Chemical elements	118	14	<Period>, <Phase-at-R.T.>

TABLE XI

THE PROMPT USED TO SELECT THE ENTITIES THAT ARE CONSISTENT WITH THE ATTRIBUTES OF S^{pos} (S^{neg}) IN THE TOP-T ENTITIES OF L_0 TO CONSTRUCT L_{pos} (L_{neg}).

I have a task that involves classifying candidate entities based on their alignment with a seed entity set. The seed entities are grouped together because they share certain attributes, referred to as seed attributes. I will provide a list of seed entities along with their seed attributes. Additionally, I have a list of candidate entities that are similar to the seed entities but may not necessarily share the same seed attributes. I need you to identify the seed attributes and use them to classify each candidate entity into one of two categories: 1) consistent with the seed entity set in terms of seed attributes, or 0) inconsistent with the seed entity set in terms of seed attributes. For the given N candidate entities, please output N values, each being 1 or 0, indicating whether each candidate is consistent (1) or inconsistent (0) with the seed entity set based on the seed attributes.

Input:
Seed entities: [Mark Twain, Ernest Hemingway, F. Scott Fitzgerald]
Candidate entities: [J.K. Rowling, Stephen King, Agatha Christie, John Steinbeck, Harper Lee, Charles Dickens, Virginia Woolf], total 7 entities
Output:
“result”: [0,1,0,1,1,0,0]

Input:
Seed entities: [Golden Retriever, German Shepherd, Labrador Retriever]
Candidate entities: [Bengal Tiger, Beagle, Siberian Husky, African Elephant, Pug], total 5 entities
Output:
“result”: [0,1,1,0,1]

Input:
Seed entities: [{Entity1}, {Entity2}, {Entity3}]
Candidate entities: [{Entity1’}, {Entity2’}, {Entity3’}, ...], total {} entities
Output:

APPENDIX F
WHY GENEXPAN OMITS NEGATIVE SEEDS IN ENTITY GENERATION

The primary reason for not using negative seeds during the entity generation stage is to maximize recall in this initial phase, ensuring the retrieval of as many entities belong to fine-grained class as possible. This intuition aligns with the first-stage retrieval (coarse ranking) in recommendation and retrieval systems.

In our preliminary experiments, we compared results with and without negative seeds, using the negative-seed-enhanced prompt described in Table XIV. As reported in Table XV, we evaluated both the first-stage recall (coarse recall in entity generation) and end-to-end model performance (ComMAP). The results show that introducing negative seeds harms recall, which in turn leads to a decline in overall MAP performance.

TABLE XII

THE PROMPT USED TO GENERATE ENTITIES THAT ARE SEMANTICALLY SIMILAR TO 3 GIVEN ENTITIES.

iron, copper, aluminum and zinc.
math, physics, chemistry and biology.
{Entity1}, {Entity2}, {Entity3} and ______

TABLE XIII

THE PROMPT USED TO GENERATE A CLASS NAME THAT COVERS THE GIVEN ENTITIES.

Generate a class name that accurately represents the following entities. This class name should encompass all the given entities and reflect their shared characteristics.
Examples:
[Tiger, Lion, Cheetah] → Big Cats
[Shakespeare, Tolstoy, Hemingway] → Famous Authors
[Mercury, Venus, Mars] → Planets in the Solar System
[{Entity1}, {Entity2}, {Entity3}] → ______

TABLE XIV

PROMPT USED IN THE GENERATION PHASE OF GENEXPAN WITH NEGATIVE SEEDS.

similar to iron, copper, aluminum, rather than wood, plastic, glass, it is zinc.
similar to math, physics, chemistry, rather than history, art, music, it is biology.
similar to {pos_ents}, rather than {neg_ents}, it is ______

TABLE XV

APPLYING NEGATIVE SEEDS TO THE ENTITY GENERATION PHASE (SIMPLY NOTED AS PHASE 1) OF GENEXPAN.

Method	**Metric**	@10	@20	@50	@100
GenExpan	Recall in Phase 1	93.58	92.43	74.14	52.30
	ComMAP	69.79	68.35	67.07	66.38
+ Neg. seeds in Phase 1	Recall in Phase 1	95.27	83.65	48.52	27.24
	ComMAP	65.91	63.79	61.05	60.02