A Appendix

A.1 Benchmark information

A.1.1 Sticker retrieval datasets. We collect two original sticker retrieval datasets: (i) WeChat [36] is a public dataset from the sticker retrieval challenge of CCIR Cup 2024 [4], with 500K stickers spanning 73 different styles and 6,250 intellectual properties (IPs, i.e., characters) from WeChat; (ii) Sticker820K [43] is another public sticker retrieval dataset constructed by other researchers containing around 820K stickers across 8 different styles. For WeChat dataset, the metadata of a sticker includes image, caption, emotion, style, IP and OCR (text recognized in the image). Here, the sticker text t comprises the caption, emotion, style, IP, and OCR and the sticker image is denoted as v. The fields of Sticker820K dataset are similar to WeChat dataset, but lack the IP field. The data examples of WeChat are shown in Figure 5.

A.2 LLM prompts

A.2.1 Prompt for CoT-based query expansion. The guiding prompt for recognizing the expressed intent is: "`Given a user query for searching stickers, {query}, it may either represent an expressive form of intent or a direct intent expression. For example, "overjoyed to the point of flying" is an expressive form, whereas "happy" is a direct intent expression. First, determine whether the query is an expressive form of intent or a direct intent expression. (i) If it is an expressive form of intent, analyze its underlying core intent and output it directly; (ii) If it is a direct intent expression, output it as is.`"

The prompt for guiding the step-by-step generation of the different expressions of the query is: "`Given a user query for searching stickers, along with its expressive intent, your task is to generate a new variant of the sticker query that aligns with the given intent while ensuring it is distinct from both the original query and any previously generated variants. The new variant should maintain the intended meaning but introduce differentiation wherever possible.`
`Query: {query}`
`Intent: {intent}`
`Pervious variants: {variants}`"

A.2.2 Prompt for generating variant sticker text expressions. The prompt for generating variant sticker text expressions is: "`Given a user query for searching stickers and the associated sticker text (including captions, text within the sticker, characters, and sticker style), the task is to identify the abstract expressive intent behind the query and the sticker text. Based on this understanding, a new variation of the sticker text should be directly generated that aligns with the given expressive intent while ensuring it is distinct from both the original text and previously generated variations. The generated variation should maintain meaningful differentiation while staying true to the intended expression.`
`Query: {query}`
`Relevant sticker text: {sticker text}`
`Pervious variants: {variants}`"

A.3 Baseline details

We compare our method with several representative approaches, including regular sticker retrieval models, retrieval methods tackling OOD queries and OOD documents (stickers, in this paper), respectively, from text & image retrieval:

- For regular sticker retrieval models, we adopt: (i) BM25 [24] is a classical probabilistic retrieval model that shows effectiveness on OOD retrieval tasks [31]. We take sticker text as the document; (ii) StickerCILP [43] directly fine-tunes CLIP model [22, 39] to capture sticker features and text features; (iii) StickerLLM [43] is similar with StickerCILP but uses ChatGLM-6B [41] as the query encoder; and (iv) Int-RA [12] matches stickers by understanding the common-sense requirements of the query.
- For retrieval methods that enhance robustness to OOD queries, we adopt: (i) DRTA [45] is from text retrieval and uses contrastive learning to bridge the gap between the query and the possible variants; (ii) DST [29] is from text retrieval and aligns the ranking list between query and variants in a self-training manner; and (iii) PlugIR [11] is from image retrieval and leverages LLM reasoning to understand and refine unseen queries.
- For retrieval methods that enhance robustness to OOD stickers, we adopt: (i) Inpars [10] is from text retrieval and generates pseudo query for OOD stickers with LLMs; (ii) COCO-DR [40] is from text retrieval and uses distributionally robust optimization to learn domain features of unseen stickers; and (iii) DAR [15] is from image retrieval and uses a diffusion model to generate unseen stickers to assist with retrieval.

A.4 Necessary Statements

The experiments were conducted on 4 × NVIDIA Tesla V100 32G GPUs. Training an XAlign-SR model takes approximately 12 hours, while evaluation requires about 1.5 hours.

We invited ethical labelers and volunteers and ensured that they received higher than the standard local hourly rate. For online test volunteers, we invited users who were proficient in using the sticker search function and ensured that all information seen by all users was fully desensitized. For the baseline, we ensured that all information was desensitized and did not reveal user or system information.

User query	Relevant sticker					
	Image	Caption	OCR	IP	Emotion	Style
你醒啦? (Are you awake?)		可爱猫醒了吗 (Is the cute cat awake?)	醒了吗 (You're up?)	动物: 猫 (Animal: Cat)	可爱 (Cute)	萌宠 拍摄 (Cute pets Shooting)
我好孤独 (I'm so lonely)		我的世界只剩下孤独 (The only thing left in my world is loneliness)	我的世界只剩下孤独 (The only thing left in my world is loneliness)	兔斯基: 饿疯兔 (Bugs bunny: Hungry crazy bunny)	日常 (Daily)	绘制表情 卡通形象 简笔画 (Drawing sticker Cartoon Sketches)
梦里啥都有 (It's all in the dream)		波吉开始做梦 (Boogie starts dreaming)	开始做梦 (Start dreaming)	国王排名: 波吉 (King's eanking: Boogie)	日常 (Daily)	动漫人物 绘制彩图 (Anime characters Drawing color)
嗨皮 (Happy (homonym))		有亿点开心 (I'm only a little (billion) happy)	有1点开心 (A little happy)	明星: 成龙(Star: Jackie Chan)	搞笑 (Funny)	真人 男 拍摄 (Real Man Shooting)

Figure 5: Data examples of query and sticker metadata in Wechat. Note that Sticker820K does not have the IP field.