

Deep Search：基于大模型的搜索技术革命

原创 kevenlw 数图笔记 2025年02月23日 22:37 上海

说完了Deep Research，再来说说更底层、更技术、更“情报”的Deep Search。这里须注意：DeepSearch不是DeekSeek，虽然也是“深度搜索”（求索？）。

搜索（检索）是情报学（信息科学）的核心概念，自从“搜索引擎”大行其道，就仿佛已成功出逃，深受左邻右里相关学科的宠爱，情报学都不好意思宣称主权。

对于搜索，情报学除了贡献了诸多情报检索理论，还发展了一整套技术应用体系，诸如情报检索语言，概念模型、规范控制方法、语义匹配规则、提问处理技术（正则表达式、同义词环等），以及各学科的词表等等，在科班出身的情报学者那里，一直有一个残念，无论技术怎样发展，都是对情报理论的具体实现，是不愿放弃对搜索进行研究的。

大模型的爆发让搜索越走越远，为搜索带来了真正的革命性技术，搜索经历着从文本和关键词匹配到真正语义理解的范式转变。

于是融合了深度学习、自然语言处理和大数据分析等前沿技术的深度搜索（Deep Search）应运而生，正在重塑人类获取和处理信息的方式，将搜索变为不需要搜索，开创智能信息服务的新纪元。

深度搜索是不是就是应用了AI的搜索？实现Deep Search的原理是什么？如何应用AI就成了Deep Search？与传统的搜索有什么区别和优势？与Deep Research有什么区别和联系？它是一种技术还是一种服务？有哪些公司推出了Deep Search服务？本文就来回

一、什么是深度搜索（Deep Search）？

深度搜索（Deep Search）是一种通过深度学习、自然语言处理（NLP）、语义分析等先进技术，提供比传统搜索更为精确、高效、智能和个性化的信息检索方式。它能够更好地理解用户的意图和语境，处理复杂或模糊的查询请求，更注重对信息的理解和语义层面的解析，而不仅仅是进行关键词匹配。

二、深度搜索Deep Search与深度研究Deep Research

虽然Deep Search与Deep Research在中文语境中均含有“深度”之意，但二者的技术定位与应用边界存在本质差异。

| 维度 | Deep Search | Deep Research |

| 目标 | 信息检索与关联挖掘 | 系统性分析与知识发现 |

| 技术重点 | NLP、知识图谱、多模态检索 | 数据建模、统计推断、因果分析 |

| 输出形式 | 精准答案、关联结果列表 | 研究报告、趋势预测、决策建议 |

| 典型场景 | 企业知识库搜索、法律案例检索 | 学术研究、市场战略分析、政策制定 |

案例对比：

Deep Search：快速找到“2023 年特斯拉电池供应商的专利技术”。

Deep Research：分析“全球电动车电池技术竞争格局，预测 2030 年市场份额”。

Deep Search作为搜索技术演进的新范式，其核心在于通过深度学习重构信息检索的全流程，强调对查询意图的深度解析与多模态数据的实时响应。而Deep Research则是建立在深度搜索技术之上的应用服务，聚焦于复杂研究任务的自动化处理与知识产品的结构化输出。

Deep Search 的本质是 **AI 驱动的语义检索革命**，其技术门槛在于多模态数据融合与动态学习能力。而 Deep Research 更侧重于深度分析与知识生成，两者结合可形成“搜索→分析→决策”的完整链路（如 ChatGPT 插件生态）。



二、深度搜索的实现原理

Deep Search 的核心是通过 **AI 技术**突破传统搜索的局限性，其技术栈包括：

| 技术模块 | 作用 |

| 自然语言处理（NLP） | 解析用户查询的语义（如识别意图、情感、实体） |

| 知识图谱 | 构建实体关联网络（如“马斯克→特斯拉→SpaceX→星链”） |

| 深度学习模型 | 使用 Transformer（如 BERT、GPT）理解上下文并生成语义向量 |

| 多模态分析 | 融合文本、图像、视频等数据（如从医学影像中提取诊断关键词） |

| 强化学习 | 根据用户点击/反馈动态优化搜索排序（如优先展示高转化率内容） |

示例流程:

当用户搜索“如何降低电动车电池成本?”时, Deep Search 会:

解析问题中的实体(电动车、电池、成本)和意图(解决方案、技术优化)。

通过知识图谱关联“锂离子电池→固态电池→宁德时代→4680 电池”。

从专利、论文、行业报告中提取非结构化数据,生成摘要式答案。

大模型相关技术的应用是实现Deep Search的前提条件。传统搜索依赖规则引擎(如关键词匹配),而Deep Search必须通过 AI 实现语义理解。例如:搜索“苹果财报”,AI 能区分“水果市场报告”和“Apple Inc. 财务数据”。

并非所有 AI 搜索都是深度搜索。AI可以仅用于优化排序(如点击率预测),就不是深度搜索。

Deep Search: 需完整覆盖 语义解析→多模态分析→动态学习 的闭环。



三、Deep Search与传统搜索的区别

理解用户意图:

- 传统搜索: 依赖用户输入的关键字进行匹配, 搜索结果是基于页面内容和关键词的相关性排序。这种方式可能在面对模糊查询或长尾关键词时效率较低。
- Deep Search: 通过深度学习和自然语言处理技术, 能够理解用户查询的语义和上下文, 准确判断用户的意图, 即便是模糊或复杂的查询也能给出合理的答案。例如, 用户查询“如何提高工作效率”, 传统搜索可能给出一堆方法列表, 而深度搜索能理解“提高效率”的背景, 返回更相关的解决方案或工具推荐。

搜索技术的基础：

- 传统搜索：基于页面内容的索引、关键词匹配和排名算法（如PageRank）。
- Deep Search：采用深度学习、语义分析、图神经网络等技术，进行更深层次的信息理解和推理。这种技术可以帮助搜索引擎超越关键词层面，识别查询背后的复杂含义。

数据源和信息类型：

- 传统搜索：主要集中在网页和外部链接的内容，且多为基于文本的检索。
- Deep Search：能够处理更多类型的数据，包括文本、图像、视频、音频等多模态信息。它不仅能检索静态的网页，还能够理解和处理多种格式的数据，提供更加丰富的搜索体验。

结果的精准度和个性化：

- 传统搜索：依赖于外部的链接和页面内容，与用户的个性化需求匹配度有限。
- Deep Search：通过个性化算法分析用户历史行为、搜索习惯等，能够更好地为每个用户定制搜索结果，提升用户体验。



Deep Search的优势

- 语义理解：传统搜索引擎往往依赖于关键词的匹配，而深度搜索能够理解查询的上下文和语义，从而更加精确地解读用户的需求。例如，“苹果”可以是水果，也可以是公司，深度搜索能根据上下文来判断用户意图。

- 个性化推荐：通过分析用户的历史行为、偏好、语言风格等，深度搜索能够为用户提供个性化的搜索结果，使得每次搜索都更加贴合个人需求。
- 多模态搜索：深度搜索不仅能处理文本信息，还能够处理图片、视频、音频等多种数据类型，支持跨平台、多种信息源的综合检索。
- 提高搜索效率：通过智能化的语义分析，Deep Search能够减少无关信息的干扰，帮助用户更快地找到真正有价值的结果，尤其在面对大规模数据时尤为显著。

四、技术架构的对照分析

（一）查询处理机制

Deep Search系统普遍采用实时语义向量化技术，属于“快思考”，如DeepSeek研发的多头潜在注意力机制，能在300毫秒内完成查询语句的语义编码。这种即时处理能力使其在电商搜索、新闻聚合等场景中占据优势。而Deep Research服务则引入强化学习框架，通过多轮迭代优化搜索策略，采用“慢思考”的推理机制才能达到最好的效果。

（二）数据交互模式

在数据吞吐层面，Deep Search强调实时数据流的处理能力。微软的架构白皮书披露，其Deep Search服务可同时处理2000万并发查询，每秒更新索引数据量达15TB。Deep Research则注重深度数据挖掘，OpenAI的技术文档表明，单个研究任务平均会扫描1200个信息源，提取有效数据点超过5000个。这种差异导致两者在硬件资源配置上呈现不同特征：Deep Search需要大规模GPU集群实现低延迟推理，而Deep Research更依赖高带宽存储系统支撑深度数据分析。



五、服务形态的差异化发展

（一）核心功能定位

Deep Search作为基础设施级服务，其功能边界聚焦于信息检索的效率优化。韩国Deep Search公司的工业级解决方案，通过边缘计算节点将工厂设备检索延迟降低至5毫秒级。Deep Research则定位于知识生产工具，如Perplexity的系统可将散乱信息整合为包含摘要、方法论、数据可视化的完整报告。这种定位差异在输出形态上表现明显：Deep Search返回结构化数据片段，而Deep Research生成的研究报告平均长度达3000字，包含8-12个数据图表。

（二）商业化路径

从商业模式观察，Deep Search多采用API调用计费模式，微软Azure的搜索服务定价为每千次查询0.15美元。Deep Research则发展出成果导向的收费体系，OpenAI的研究服务按报告复杂度收费，基础版每月199美元包含50次研究任务。这种差异反映出两者不同的价值创造逻辑：前者作为信息管道按流量收费，后者作为知识工厂按产出定价。当然这两者现在都面临激烈的竞争和开源的威胁，只要技术无法垄断，总是会向低收费的公共服务模式的方向发展。

六、应用场景的互补与竞争

（一）时效性需求场景

在金融交易、舆情监控等对实时性要求苛刻的领域，Deep Search展现不可替代性。高频交易系统的测试数据显示，Deep Search的毫秒级响应使套利机会捕获率提升15%。而Deep Research在此类场景中的应用多限于事后分析，如Perplexity的系统需至少3分钟生成市场异动分析报告。

（二）知识密集型场景

对于学术研究、政策分析等需要深度知识加工的场景，Deep Research的价值凸显。在生物医学领域，OpenAI的服务可将文献综述时间从40小时压缩至2小时，同时引文准确率达到98%。Deep Search在此类场景中更多承担初步信息筛选功能，无法替代深度研究的知识合成过程。

七、技术演进的融合趋势

前沿技术发展正在模糊Deep Search与Deep Research的边界：



微软研究院的Project Galileo尝试在搜索流程中嵌入研究型AI代理，使单次查询即可触发自动化的深度分析

DeepSeek开发的MoE-12B模型，将搜索延迟与研究深度均衡在1秒响应+85%分析深度的新平衡点

Perplexity正在测试的流式研究报告生成技术，可将30分钟的研究任务压缩至5分钟内逐步输出

这种融合趋势在硬件层面催生新型计算架构，如谷歌最新发布的TPU v6处理器，其混合精度计算单元既可支撑低延迟搜索，又能加速深度研究任务，使两类服务的运行成本降低40%。

Deep Search与Deep Research共同构成智能信息服务的双重引擎，前者革新信息获取效率，后者重塑知识生产模式。技术供应商需要根据应用场景的实时性需求、知识密度要求进行精准定位。未来三年，随着多模态大模型与边缘计算的发展，两类服务将呈现"底层架构融合，应用层分化"的演进态势。企业用户应当建立搜索-研究协同 workflow，在即时决策与深度分析间实现动态平衡，而个人用户则需要培养新型信息素养，以有效驾驭这两种智能化程度不断提升的工具。

