



多语言软件扩展解决方案

17 种新增语言

25 个项目

从零开始的组件实施

本案例中的客户需要一家合作伙伴来帮助他们进行多语言软件实施框架扩展。我们帮助其实现了这一目标 — 在 25 个项目中提供了 17 种新增语言。

面临的挑战

本案例中的客户建立了一个能够支持多语言的软件实施框架。该框架依赖于有监管和无监管的机器学习方法以及文本分析、预处理和后处理程序，包括词语切分和词性标注。

该客户开发了一套工具和流程用于将工作组件实施为数个语言版本，并完成了相关测试。但客户希望寻找一位合作伙伴来记录、开发和维护该框架，并将组件实施为更多语言版本。

客户简介

本案例中的客户是全球大型企业软件提供商，是自然语言文本技术领域的领军者。该客户构建了一整套产品，其中包括一套高效算法，经过该算法处理的文本可供其他应用程序使用。



关键术语

命名实体识别器

在给定文本中，找出命名实体并确定它们属于哪些预定义类别（例如，人名、地名）。

例如，Peter 是一名工程师，住在多伦多 Peter = 人，多伦多 = 地点

关键词语提取器

根据与给定文本上下文之间的相关性提取关键词语或关键词列表。

例如，对于上面这句话，提取的列表是：1. 关键词语提取器 2. 关键词语列表 3. 给定文本

成果

Lionbridge 针对 17 种新增语言高效地帮助客户开发了扩展解决方案。对于其中大多数语言，软件的实施都是从零开始的。我们创建了所有必要的文档、路线图和辅助工具，以便使用面向任意语言的客户框架从零开始实施新软件并改进现有软件。我们还开发了一套体系，以记录为使该软件达到令人满意的性能水平所采取的所有步骤。最后，我们向客户咨询了如何进一步改进框架及其组件。

我们成功在 25 个项目中达成了这些成就，而且我们的客户已将最终解决方案集成到了基于云的 API 中，从而提供了基于基本文本的自然语言处理服务。

更多详情，请访问

LIONBRIDGE.COM

解决方案

Lionbridge 针对 17 种新增语言高效地帮助客户开发了扩展解决方案。对于其中大多数语言，软件的实施都是从零开始的。



1. 建立高效的工作流程

在合作过程中，Lionbridge 先从客户处获得指导，在这期间，我们使用客户的框架启动了一种试点语言的软件实施。然后我们在内部继续进行试点语言的软件实施，并同步展开另外几种语言的软件实施。这种做法确保了及时执行并实现了流程改进。



2. 克服特定语言挑战

在处理不同语言的过程中，我们的团队能够成功应对特定语言的挑战，包括：

命名实体识别：大写挑战

- 有些语言（例如，阿拉伯语、日语、韩语）没有大写形式，而大写形式是命名实体识别中的一大主要特征。

关键词语提取：过度生成

- 提取关键词语的依据是，它包含某个在文本中经常出现的特定词语；但能否正确提取，存在一定的不确定性。如果提取出的关键词语同样包含特定高频词，但与给定文本无关，就属于错误提取。

Lionbridge 团队执行了基于语言的规则编写脚本，我们将这些脚本集成到框架中，以应对跨语言的挑战。我们的团队还实施了文本处理辅助工具来支持软件开发，并运用了多种分析工具来轻松确定需要改进之处。



3. 提供坚实的后盾支持

Lionbridge 建立了一支庞大的自然语言处理团队，该团队由语言学家、项目经理、数据工程师和全球语言专家网络组成，致力于创建面向几乎所有语言的自然语言处理解决方案和资产。我们部署了基于云的平台，让我们的全球团队能够使用 Lionbridge 的工具在任何地点工作，并获得 Lionbridge 项目经理和协调员的指导和监督。我们的业务范围遍布全球，因而我们的团队可以随时随地为我们的客户提供高效服务。

我们的客户需要一家能够提供远超传统本地化项目所需服务的公司。它需要一个深刻了解机器学习方法的合作伙伴，该合作伙伴应可以开发并增强文本预处理和后处理程序以及基于语言的规则编写。普通的翻译人员并不具备这些技能背景。Lionbridge 培养了具备机器学习知识的专家网络，并部署了相关设施，服务范围涵盖 200 多种语言。