### **PHP Googlebot 识别与链接管理系统 - 开发需求文档**

#### **文档信息**

* ****版本****：1.0
* ****状态****：初稿
* ****最后更新日期****：2025 年 07 月 29 日

#### **1. 引言**

##### **1.1 文档目的**

本文档旨在明确「PHP Googlebot 识别与链接管理系统」的功能需求、非功能需求及技术规范，为开发、测试及维护提供依据。

##### **1.2 项目背景**

为实现对网站首页及一级目录页面的智能管理，需通过程序识别 Google 爬虫（Googlebot），并根据服务器交互结果动态插入指定链接。同时，为避免 API 地址被封，需支持多地址切换及心跳机制。

##### **1.3 范围**

* ****适用场景****：网站首页及一级目录页面的 Googlebot 识别与链接动态插入。
* ****核心目标****：通过 API 交互维护 Googlebot IP 列表，根据访问 IP 判断是否为 Googlebot，进而动态输出指定链接，并保障 API 通信稳定性。

#### **2. 功能需求**

##### **2.1 API 地址管理**

* 初始化多组 API 地址（含 1 个主地址、N 个备用地址），支持动态更新。
* 接收服务器返回的最新 API 地址列表后，自动覆盖本地配置（用于防止地址被封）。

##### **2.2 页面判断与请求触发**

* 仅对「网站首页」或「一级目录页面」生效。
* 读取配置文件判断是否符合跳转要求：
  + 若符合且当前时间与上次请求间隔超过「心跳包间隔时间」（默认 30 分钟），触发请求；
  + 支持通过updateSD.php?AC=DO强制执行更新（忽略间隔时间）。
* 请求方式：优先异步请求，避免阻塞页面响应。

##### **2.3 数据交互规范**

###### **2.3.1 本地程序与updateSD.php交互**

* updateSD.php为独立程序，仅与本地数据交互，负责将本地临时存储的新增 Googlebot IP 提交至主服务器 API（http://www.a.com/api）。
* updateSD.php参数说明：
  + AC=数字：指定心跳包间隔时间（如AC=30代表间隔 30 分钟以上才更新）；
  + AC=DO：强制执行更新，忽略间隔时间。

###### **2.3.2 发送请求数据格式**

请求数据包需包含：

* 新增的真实 Googlebot IP 列表（本地临时存储的新 IP）。

###### **2.3.3 接收服务器返回数据格式**

服务器返回 JSON 数据需包含：

* googlebot\_ips：最新的 Googlebot IP 列表（如[0.0.0.0, 1.1.1.1]）；
* links：主数据链接及锚文本列表（如[{a: "http://www.baidu.com", b: "百度"}]）；
* insert\_dir：是否插入目录页的标志（布尔值，默认false）；
* insert\_content：是否插入内容页的标志（布尔值，默认false）；
* api\_list：最新的 API 地址列表（如['http://www.a.com/api', ...]）；
* heartbeat\_interval：心跳包间隔时间（单位：分钟，默认 30）；
* updateTime：数据更新时间（格式：Y-m-d H:i:s，如2022-12-12 16:16:00）。

##### **2.4 本地缓存管理**

* 接收服务器返回数据后，自动覆盖本地缓存（包括 Googlebot IP 列表、链接数据、API 地址、心跳间隔等）。
* 本地临时存储新增的 Googlebot IP，待updateSD.php提交后清空。

##### **2.5 链接插入逻辑**

* 触发条件：
  1. 页面在配置文件指定的「需插入链接范围」内；
  2. 访问 IP 在 Googlebot IP 列表中。
* 输出规则：链接 HTML 直接在程序代码所在位置输出（支持通过变量传递调整位置，如友链区、导航栏等）。

##### **2.6 Googlebot 识别机制**

* 基础判断：访问 IP 在本地 Googlebot IP 列表中 → 直接判定为 Googlebot。
* 补充验证（IP 不在列表但 User-Agent 为 Googlebot 时）：
  + 执行nslookup [IP]命令查询 IP 的 DNS 信息；
  + 若 DNS 解析结果包含googlebot.com相关域名 → 判定为 Googlebot。
* 结果处理：若判定为 Googlebot，将 IP 加入本地 Googlebot IP 列表及「新增 IP 列表」，并触发链接输出。

#### **3. 非功能需求**

##### **3.1 安全性**

* 对输入 IP、API 返回数据进行验证与过滤，防止恶意数据注入。
* 限制updateSD.php的访问权限，避免未授权调用。

##### **3.2 性能**

* 异步请求机制确保页面响应时间不受 API 交互影响（响应延迟≤100ms）。
* 本地缓存减少重复 IO 操作，缓存数据有效期与心跳间隔保持一致。

##### **3.3 可扩展性**

* 模块化设计（如 API 管理、IP 验证、链接输出等独立模块），支持新增功能（如多爬虫识别、自定义插入规则）。

##### **3.4 可维护性**

* 代码注释覆盖率≥80%，关键逻辑附详细说明。
* 配置文件集中管理（如 API 地址、心跳间隔等），便于修改。

#### **4. 总结**

本系统通过 API 动态维护 Googlebot IP 列表，结合本地验证机制精准识别 Google 爬虫，实现链接的动态输出，并通过多 API 地址切换及心跳机制保障稳定性，适用于需针对 Googlebot 优化的网站场景。