

【발명의 설명】

【발명의 명칭】

JSON Lines(JSONL) 포맷 기반의 사전 압축 및 복원 방법{Method for Dictionary Compression and Restoration based on JSON Lines (JSONL) Format}

【기술분야】

본 발명은 데이터 압축 및 복원 기술에 관한 것으로서, 특히 JSON Lines (JSONL) 포맷을 대상으로 하는 스트리밍 기반 사전(Dictionary) 압축 및 복원 방법 및 이를 구현하는 시스템/매체에 관한 것이다. 본 발명은 AI 학습 데이터 파일과 인, 로그/텔레메트리 스트림, IoT 데이터 수집 등에서의 대규모 텍스트 데이터 전송·저장 최적화와, 압축 후에도 가독성과 디버깅 가능성을 유지하는 JSONL 포맷 호환성 기술에 관한 것이다.

【발명의 배경이 되는 기술】

JSON 포맷은 구조적 가독성이 높아 광범위하게 사용되지만, 전체 문서의 트리 구조 특성상 스트리밍 처리 및 부분 복원이 어려워 대용량 환경에서 메모리와 시간 비용이 크다.

종래, JSON 포맷 데이터의 효율적인 압축 및 복원 기술에 관한 대한민국 공개특허공보 제10-2023-0015776호 (이하, "선행기술 1")가 알려져 있다. 선행기술 1은 텍스트 형태를 유지하면서 압축률을 높이는 JSON 포맷 압축/복원 방법을 제시하였으며, 본 발명의 일부 발명자는 선행기술 1의 공동 발명자로 참여하였고, 해당 발명을 통해 일정 수준의 압축 효율을 달성하였다. 그러나, 선행기술 1은 대규모 스트

리밍 환경 (예: LLM 학습 데이터셋, 실시간 로그 스트림 등)에 적용될 경우, 전체 데이터를 로드하여 처리 및 복원해야 하는 구조적 한계를 가진다.

라인 독립성 기반 스트리밍 최적화 미흡: JSONL 포맷의 핵심 특성인 라인 독립성을 활용한 스트리밍 최적화가 실질적으로 구현되지 않아, 대규모 데이터 처리 시 전체 문서 메모리 적재 또는 높은 메모리 오버헤드가 발생한다.

메모리 비의존적 운용 모델 부재: JSONL 스트림 처리 시 입력 크기에 비례하지 않는 일정한 메모리 상한에서 동작하는 것이 필수적이나, 선행기술 1은 이러한 실질적 운용이 어렵다. 따라서 본 발명자는 선행기술 1의 한계를 극복하고, JSONL의 특성을 직접 활용하여 (1) JSONL 포맷을 유지한 채, (2) 전체 파일 적재 없이, (3) 라인별 사전 관리와 세그먼트 동기화로, (4) 실시간 스트리밍에서 안정적으로 압축 · 복원이 가능한 새로운 구조를 고안하게 되었다.

【선행기술문헌】

【특허문헌】

대한민국 공개특허공보 제10-2023-0015776호 (JSON 포맷 압축 및 복원 방법)

【발명의 내용】

【해결하고자 하는 과제】

본 발명은 상기 선행기술 1의 구조적 한계를 극복하고, JSONL 스트리밍 환경에 최적화된 압축/복원 방법을 제공하는 것을 목적으로 한다. 구체적 목적은 다음과 같다.

선행기술 1의 메모리 의존성 문제 해결: 대용량 JSONL 데이터에 대하여, 전체 파일 메모리 적재 없이 일정한 메모리 상한에서 스트리밍으로 처리하면서 높은 압축률을 달성한다.

JSONL 포맷 호환성 및 운영 안정성 확보: 압축 후에도 JSONL 포맷을 유지하여 범용 파서로 즉시 구문해석 가능하게 하며, 특히 선행기술 1에서 미흡했던 대규모 스트림 환경에서의 라인 독립적 디버깅 가능성 및 운영 안정성(오류 격리 및 부분 복원 기능)을 확보한다.

라인별 사전 관리 체계 확립: 선행기술 1에서 구체적으로 제시되지 않았던 라인별 사전 관리와 세그먼트 기반 버전 관리를 통해 부분 복원 및 오류 격리를 실현한다.

【과제의 해결 수단】

본 발명은 JSONL 포맷 호환성을 유지하면서 반복되는 키 · 값 · 키세트를 사전 기반으로 압축 · 복원하는 방법을 제안한다. 주요 특징은 다음과 같다.

반복되는 키 · 값 · 키세트를 대응코드(kN/vN/sN)로 치환한다.

사전은 DIC-JSONL 구조(사전 라인으로서의 JSON 객체)로 직렬화되어 데이터 내 독립 라인으로 삽입되며, 범용 JSONL 파서로 그대로 읽힐 수 있다.

라인 독립성에 기반하여 오류를 라인 단위로 격리하며, 필요 시 세그먼트 경계를 정의하여 복수 사전 운용 및 버전 관리를 적용한다.

긴 스트리밍 환경에서는 SYNC 마커를 선택적으로 삽입하여 오류 격리 및 재동기화를 지원할 수 있다.

【발명의 효과】

본 발명은 다음과 같은 효과가 있다.

JSONL 포맷 유지로 가독성과 범용 파서 호환성을 확보하며, 디버깅과 운영성이 크게 향상된다.

스트리밍 방식 및 일괄 처리 방식 모두를 지원하여, 소규모 실시간 처리부터 대용량 데이터 처리까지 유연하게 대응 가능하다.

라인별 사전 관리와 라인 독립성에 기반한 오류 격리로 부분 복원 및 병렬 복원이 가능하다.

선택적인 동기화 마커(SYNC) 적용 시, 오류 발생 위험이 높은 스트리밍 환경에서 장애 복구와 재처리 비용이 더욱 감소한다.

【도면의 간단한 설명】

도 1은 JSONL 스트리밍 사전 압축 및 복원 전반 흐름을 도시한다.

도 2는 입력 전처리 및 오류 큐 분기 흐름을 도시한다.

도 3은 사전 기반 압축 절차를 도시한다.

도 4은 사전 기반 복원 절차를 도시한다.

도 5는 JSONL 원본 데이터 예시를 도시한다.

도 6은 DIC-JSONL 사전 및 압축 라인의 예시를 도시한다.

【발명을 실시하기 위한 구체적인 내용】

1. 시스템 구성 및 동작 개요

시스템은 포맷추출부, 포맷서치부, 사전관리부, 포맷압축부, 포맷복원부로

구성되며, 선택적으로 동기화부를 포함할 수 있다. 각 부는 라인 경계 단위의 독립 처리 원칙을 따른다.

입력은 개행으로 분리된 JSONL 스트림이며, 각 라인은 독립적 JSON 객체이다. 출력은 압축된 JSONL 라인과 주기적으로 삽입되는 사전 DIC-JSONL로 이루어지며, 필요 시 SYNC 마커를 추가할 수 있다.

사전 DIC-JSONL은 표준 JSON 객체 형식으로 직렬화되어 범용 JSONL 파서가 읽을 수 있다.

사전 라인과 일반 데이터 라인을 구별하기 위해 식별 필드(예: "_type": "dict_header" 또는 "_dict": true)를 포함하면, 파서는 이 필드의 존재 여부로 해당 라인이 사전인지 데이터인지를 판단한다.

예를 들어, {"_type": "dict_header", "dict": {"k": {"userId": "k11", ...}, "v": {"OK": "v03", ...}, "s": {[{"name": "age"}]: "s2"}}, "ver": 7, "hash": "..."}와 같이 표현되며, 본 발명은 특정 키명을 한정하지 않고 JSON 유효성과 범용 파서 호환성을 우선한다.

사전은 세그먼트 경계에서만 버전이 증가한다. 라인 처리 도중 버전 변경은 허용하지 않음으로써 라인 일관성을 보장한다.

2. 처리 방식과 메모리 운용

본 발명은 일괄 처리 방식 및 스트리밍 처리 방식 모두를 지원한다. 일괄 처리 시에는 전체 데이터를 메모리에 적재하여 처리할 수 있으며, 스트리밍 처리 시에는 라인 단위로 순차 처리하여 메모리 사용량을 제한할 수 있다.

스트리밍 처리 모드에서는 사전 크기와 버퍼 크기를 제한하여 일정한 메모리 상한을 유지할 수 있다. 이 경우 메모리 사용량은 입력 크기와 무관하게 일정 범위 내로 제한된다.

복수 세그먼트 운용 시, 세그먼트 경계에서 LRU 또는 빈도 기반 정책으로 사전 엔트리를 교체할 수 있으며, 사전 교체는 오직 세그먼트 경계에서만 수행되어 라인별 복원 일관성이 유지된다. 단일 사전 운용 시에는 교체 없이 고정 사전을 유지한다.

3. 라인별 사전 관리와 동기화

라인별 사전 관리는 운용 모드에 따라 다음과 같이 구현된다.

- 단일 사전 모드: 전체 스트림에 대해 하나의 사전 집합을 사용하며, 버전 관리가 불필요하다. 각 라인은 동일한 사전을 참조하므로 별도의 버전 표기나 인덱싱 없이 압축/복원이 가능하다.

- 복수 사전 모드: 여러 세그먼트에 대해 각기 다른 사전을 운용하는 경우, 다음 두 방식 중 하나 또는 복합으로 사전 관리를 구현한다:

- 내장식 표기: 각 라인에 최소 메타 필드(예: 사전 버전 값)를 포함시켜 해당 라인의 참조 사전을 명시.
 - 외장식 인덱스: 세그먼트 헤더에 사전 버전을 기록하고, 라인 오프셋 범위와 버전을 매핑하는 인덱스를 별도로 유지.

동기화 마커(SYNC)는 선택적 구성요소로서, 긴 스트리밍이나 오류 격리가 중요한 환경에서 사용된다.

SYNC를 사용하는 경우, 사전 식별 정보와 엔트리 수를 포함하며, 일정 라인 간격 또는 사전 갱신 시 삽입된다.

오류 발생 시 최근 SYNC 이후 라인만 재처리하여 재처리 비용을 절감한다.

SYNC에 선택적으로 사전 DIC-JSONL의 해시/체크섬이나 최근 변경 요약을 포함하여 무결성 검증을 수행할 수 있으나, 이는 필수 요소가 아니다.

단일 세그먼트 또는 실시간 학습 환경에서는 SYNC 마커를 생략할 수 있다.

라인 독립성에 기반하여 다음의 부분 처리 모드를 지원한다.

- 범위 복원: 지정 라인 범위만 복원.
- 실시간 복원: 도착 순서대로 즉시 복원.
- 병렬 복원: 라인 단위 태스크를 병렬로 처리(라인별 참조 사전이 명확하므로 경합이 없음).

4. 사전 엔트리 설정과 코드 치환

사전 엔트리 설정은 임계치와 길이 조건을 만족하는 반복 토큰에 한정한다. 지나치게 짧은 토큰(예: 길이 1~2)은 제외하여 오버헤드를 줄인다.

키세트는 동일 순서의 키 목록 반복을 검출하여 sN 코드로 등록한다. 키세트 치환 시 값은 배열로 직렬화되어 라인 길이를 대폭 단축시킨다.

치환 우선순위는 키세트 → 키 → 값이다. 이는 구조 보존과 구문 안정성을 위한 규칙으로, 서로 다른 범주의 코드가 중첩될 때 일관된 결과를 보장한다.

대응코드는 원문과 충돌하지 않도록 예약 접두 규칙("k", "v", "s" 접두어와 숫자 조합) 및 코드 공간 확장을 사용한다.

충돌 회피 알고리즘은 다음과 같이 작동한다:

- (1) 대응코드 생성 시(예: k99) 원문 데이터에 동일 문자열 "k99" 키 또는 값이 존재하는지 검사,
- (2) 충돌 발견 시 숫자 증가(k99 → k100) 또는 구분자 삽입(k99 → k_99)으로 새 코드 생성,
- (3) 새 코드가 여전히 충돌하면 코드 공간 확장(예: kk99, k1_99 등) 또는 접두어 변경(예: kx99),
- (4) 충돌이 해소될 때까지 반복하여 유일한 코드를 확보. 이 논리는 세그먼트 사전 구축 시점에 수행되며, 일단 확정된 코드는 해당 세그먼트 내에서 일관되게 유지된다.

5. DIC-JSONL 사전과 포맷 호환성

사전은 JSON 표준 구문을 따르는 DIC-JSONL으로 직렬화된다. 본 사전은 다음의 논리 영역을 포함한다.

- 키 사전: 원문 키 → kN 맵핑
- 값 사전: 원문 값(문자열 등) → vN 맵핑
- 키세트 사전: 키 배열 → sN 맵핑
- 선택적 정보: 엔트리 수(entryCount), 버전(ver, 복수 사전 운용 시), 해시/체크섬(hash/checksum, 무결성 검증 시), 서명(signature, 보안 강화 시)

DIC-JSONL 사전은 JSON 스트림 내에서 독립 라인으로 존재하므로, 범용 JSON 파서로 읽기/로깅/디버깅이 가능하다.

본 발명은 사전 라인과 일반 데이터 라인을 구분하는 식별 필드(예: "_type": "dict_header", "_dict": true, "_meta": "dictionary" 등)를 포함하도록 설계된다.

파싱 규칙은 다음과 같다: (1) 각 라인을 JSON 객체로 파싱, (2) 미리 정의된 식별 필드 존재 여부 검사, (3) 식별 필드가 있으면 사전 라인으로 분류하여 사전 로딩 처리, (4) 식별 필드가 없으면 일반 데이터 라인으로 처리.

식별 필드의 구체적 명칭은 구현에 따라 달라질 수 있으나, 범용 JSON 파서가 표준 JSON 구문으로 읽을 수 있어야 하며, 사전과 데이터를 명확히 구별할 수 있어야 한다.

6. 복원 절차

복원 시 수신자는 스트림을 순차적으로 읽어 사전 DIC-JSONL을 우선 로드한다. 사전 버전에 대응하는 매핑이 준비되면, 이후 라인의 kN/vN/sN 코드를 원문으로 역치환한다.

키세트 복원은 사전에 등록된 키 배열 순서를 따라 값 배열을 매핑한다. 배열 길이 불일치가 검출되면 해당 라인을 부분 복원으로 마킹하고 오류 로그를 남긴다.

라인별 사전 참조 식별자 또는 인덱스가 버전 불일치를 방지하므로, 특정 범위만 선택적으로 복원하는 것이 가능하다.

7. 오류 처리 및 선택적 보안 기능

사전에 없는 코드가 발견될 경우, 해당 라인만 원본 형태로 패스하거나 오류

큐에 적재한다. 라인 독립성에 기반하여 오류를 라인 단위로 격리한다.

SYNC 마커를 사용하는 경우, 오류 발생 시 최근 SYNC 이후 라인만을 재처리하여 재처리 비용을 절감할 수 있다.

선택적 보안 기능으로서 다음을 포함할 수 있다:

- 무결성 검증: 사전 라인 또는 세그먼트 단위에 대해 SHA-256, CRC32 등 해시/체크섬 알고리즘 적용 가능. 검증 범위는 (a) 사전 라인 단독, (b) 사전+SYNC 루음, 또는 (c) 세그먼트 전체 중 선택.
- 서명 처리: 위 · 변조 방지 및 감사 · 포렌식 목적으로 서명 메타 필드 포함 가능.

이러한 보안 기능은 필수 요소가 아니며, 데이터 크기 증가 우려 시 생략 가능하다.

8. 성능/압축률 지표와 평가

압축률은 다음과 같이 정의된다.

$$\text{압축율} = 1 - \frac{\text{사전크기} + \text{압축된크기}}{\text{원본크기}}$$

여기서, 원본크기는 원본 총 길이, 압축된크기는 치환된 라인 총 길이, 사전크기는 사전(DIC-JSONL 및 SYNC 포함)의 총 길이이다.

세그먼트 크기, 임계치, 길이 필터, 코드 공간 정책을 조정하여 압축률과 처리 자연의 균형을 맞춘다.

9. 실시예

9.1. 기본 실시예

실시예 1: 대용량 전자상거래 제품 데이터 압축 및 복원

상황:

Kaggle Amazon Product Dataset (amz_products_small.json)에서 추출한 100,000개의 전자상거래 제품 정보로 구성된 대규모 JSONL 데이터(141.87 MB).

각 제품은 {"also_buy", "also_view", "asin", "brand", "category", "description", "feature", "image", "price", "title", "main_cat"} 11개 키로 구성.

"category" 배열 내부에 "Automotive"(9,550회), "Replacement Parts"(4,503회), "Exterior Accessories"(1,478회) 등 카테고리명이 높은 빈도로 반복 출현.

"description", "feature" 배열 내부에도 표준화된 문구가 반복. "brand", "main_cat", "price" 등 최상위 값도 데이터 전반에 걸쳐 반복 출현.

조치:

사전 등록 임계치 최소 등장 빈도 = 2, 최소 길이 = 2 설정.

전체 11개 키를 키 사전에 등록(k1~k11).

배열 및 중첩 객체 내부까지 재귀적으로 탐색하여 반복 출현하는 215,182개의 값을 값 사전에 등록(v1~v215182: 카테고리명, 브랜드명, 표준 문구, 가격 등).

동일 키 세트(11개 키) 반복을 검출하여 s1으로 등록 후 값 배열로 직렬화.

압축 시 배열 내부 각 요소를 재귀적으로 사전 참조 코드로 치환, 복원 시 역방향 재귀 치환 수행. 압축 처리 속도 8.47 MB/s, 복원 처리 속도 5.53 MB/s 달성.

원본 파일 SHA-256 해시(8de58e024a36c30f)와 복원 파일 해시 완전 일치 검증.

제품당 평균 처리 시간 0.335ms, 초당 2,985개 제품 처리 성능 확인.

효과:

대규모 데이터(100,000개 제품, 141.87 MB)에서 압축 파일 크기 92.53 MB, 49.34 MB 절감(압축률 34.78%)을 달성.

배열 내부 값의 재귀 압축으로 최상위 값만 압축 시(11.25%) 대비 약 3.1배 압축률 향상 달성.

총 처리 시간 33.50초(압축 16.75초 + 복원 16.75초)로 실시간 스트리밍 환경에서 충분한 처리 성능 입증.

복원 후 원본과 데이터 무결성 100% 일치(SHA-256 해시 검증), 배열 순서 및 중첩 구조 완벽 보존, 키 순서 유지, JSON 제어 문자(\u007f 등) 이스케이프 정확성 확인.

라인 독립성 유지로 부분 복원 및 병렬 처리 가능성 검증.

JSON 포맷 완전 호환으로 범용 파서 즉시 구문해석 가능.

값 사전 크기(215,182개)가 크지만 전체 데이터(100,000개 × 평균 11개 필드 = 약 110만 필드) 대비 약 19.5% 수준으로 중복 제거 효과 뛰어남.

전자상거래 제품 카탈로그, 재고 관리 시스템, 가격 비교 서비스, LLM 학습 데이터셋 등 실무 응용 환경에서의 즉시 적용 가능성 확인.

【산업상 이용가능성】

본 발명은 다음 분야에서 직접 적용 가능하다.

- AI/ML 파이프라인: 대규모 학습 데이터의 저장 · 전송 최적화와 실시간 데이터로더 호환.
 - 로그/보안 관제: 실시간 로그 스트리밍의 대역폭 절감과 오류 구간 신속 재처리.
 - IoT/엣지 컴퓨팅: 제한된 메모리/대역폭 환경에서의 경량 전송.
 - 클라우드 API: 반복 구조 응답의 효율적 전송과 클라이언트 복원.
- 텍스트 호환성 유지 특성은 디버깅 · 감사 · 포렌식에서 특히 유용하다.

【부호의 설명】

100 : JSONL 포맷 압축 모듈 110 : JSONL 원천 데이터

120 : 포맷추출부 130 : 포맷서치부

140 : 포맷압축부 150 : 압축된 JSONL 데이터

200 : JSONL 포맷 복원 모듈 210 : 압축된 JSONL 데이터

220 : 포맷복원부 230 : 원본 JSON 데이터 복원

【청구범위】

【청구항 1】

JSON Lines(JSONL) 포맷으로 구성된 데이터를 대상으로, JSONL의 라인 독립성 및 파서 호환성을 유지한 채 사전 기반 압축과 복원을 수행하는 방법으로서,

(a) 입력 데이터로부터 JSON 객체를 라인 단위로 추출하되, 스트리밍 방식 또는 일괄 처리 방식으로 처리 가능한 단계;

(b) 상기 추출된 라인들의 순차 처리 과정에서, 선택적으로 정의되는 세그먼트 경계(라인 수 또는 바이트량 기준) 내에서 반복 출현하는 키, 값 및 동일 순서의 키 집합(키세트)을 검출하여 사전(Dictionary)을 구축하되, 세그먼트 경계 사용 시 바이트량 기준에서 경계에 걸치는 라인은 다음 라인 시작점까지 현재 세그먼트에 포함시켜 라인 독립성을 유지하는 단계;

(c) 상기 사전을 JSONL의 텍스트 호환성을 유지하는 DIC-JSONL 구조(사전 라인 또는 세그먼트 헤더로서의 JSON 객체)로 직렬화하여 스트림에 삽입하되, 복수의 세그먼트가 운용되는 경우에는 각 사전에 고유 버전을 부여하고, 단일 사전 집합 운용 시에는 버전 없이 모든 라인이 동일 사전을 참조하도록 하는 단계;

(d) 각 데이터 라인에서 반복되는 키, 값 및 키세트를 대응코드(키: kN, 값: vN, 키세트: sN 형식)로 치환하되, 치환 결과가 단일 라인 JSON 객체로서 유효한 구문을 유지하도록 JSONL 포맷을 보존하는 단계;

(e) 복수의 사전이 존재하는 경우, 상기 치환된 각 라인이 복원 시 필요한 유효한 사전을 특정할 수 있도록 각 라인에 사전 참조 식별자를 포함시키거나 세그

먼트 헤더와 라인 오프셋을 매핑하는 인덱스 정보를 유지하며, 단일 사전 운용 시에 는 이를 생략할 수 있는 단계;

(f) 압축 데이터의 복원 시, 스트림에서 사전 DIC-JSONL을 로드하여 라인별 참조 식별자 또는 인덱스를 근거로 해당 라인의 대응코드를 원문 키, 값 및 키세트로 역치환하여 원래의 JSONL 라인으로 복원하는 단계를 포함하고,
상기 (b) 내지 (f)의 일련의 처리에서 라인 경계 단위로 오류를 격리하는 것을 특징으로 하는 JSONL 사전 압축 및 복원 방법.

【청구항 2】

제1항에 따른 방법을 수행하는 시스템으로서,
(a) 입력 스트림으로부터 JSON 객체를 라인 단위로 추출하는 포맷추출부;
(b) 반복 토큰을 탐지하고, 임계치 · 길이 필터에 따라 사전을 구축하며, 복수 세그먼트 운용 시 버전을 관리하는 포맷서치부 및 사전관리부;
(c) 각 라인의 치환 우선순위(키세트 패턴)에 따라 JSONL 포맷을 유지하면서 대응코드로 치환하고, DIC-JSONL 사전을 삽입하는 포맷압축부;
(d) 사전 DIC-JSONL을 라인 경계에서만 메모리에 로드하고, 라인별 사전 참조 식별자 또는 단일 사전에 기초하여 복원(부분 복원 · 병렬 복원 · 실시간 복원)을 수행하는 포맷복원부;를 포함하며, 대응코드가 원문 데이터와 충돌하지 않도록 예약 접두 규칙(예: "k", "v", "s"의 식별 범주와 숫자 조합)을 따르고, 복수 세그먼트 운용 시 사전 엔트리가 LRU 또는 빈도 기반 정책으로 세그먼트 경계에서만 교체되도록 구현되는 것을 특징으로 하는 JSONL 사전 압축 및 복원 시스템.

【청구항 3】

제1항 또는 제2항에 있어서,

상기 방법 또는 시스템은 선택적인 부가 기능으로서,

사전 식별 정보 및 엔트리 수를 포함하는 동기화 마커(SYNC)를 삽입하여 긴 스트리밍 환경에서 오류 발생 시 최근 동기화 지점 이후 라인만 재처리할 수 있도록 하며,

사전 라인 또는 세그먼트 단위에 대해 SHA-256, CRC32 등 표준 해시 또는 체크섬 알고리즘을 적용하여 무결성을 검증하고,
위 · 변조 방지 및 감사 · 포렌식 목적으로 사전 라인 또는 SYNC 마커에 서명 메타 필드를 포함할 수 있는 것을 특징으로 하되,

상기 기능들은 필수 요소가 아니며 운용 환경의 요구사항 및 데이터 크기 증가 우려에 따라 선택적으로 적용하거나 생략 가능한 JSONL 사전 압축 및 복원 방법 또는 시스템.

【요약서】

【요약】

본 발명은 JSONL 포맷 데이터를 대상으로 JSONL 포맷 호환성을 유지하면서 반복되는 키·값·키세트를 사전 기반으로 압축·복원하는 방법을 제공한다.

해결 수단:

- 반복되는 키·값·키세트를 검출하여 사전에 등록하고, 이를 대응코드 ($kN/vN/sN$)로 치환한다.
 - 사전은 DIC-JSONL 구조(사전 라인으로서의 JSON 객체)로 직렬화되어 데이터 내 독립 라인으로 삽입되며, 범용 JSONL 파서로 그대로 읽을 수 있다.
 - 복원 시 사전을 로드하여 대응코드를 원문으로 역치환한다.
 - 각 라인은 압축 전후 모두 유효한 JSON 객체로 유지된다.

선택적 기능:

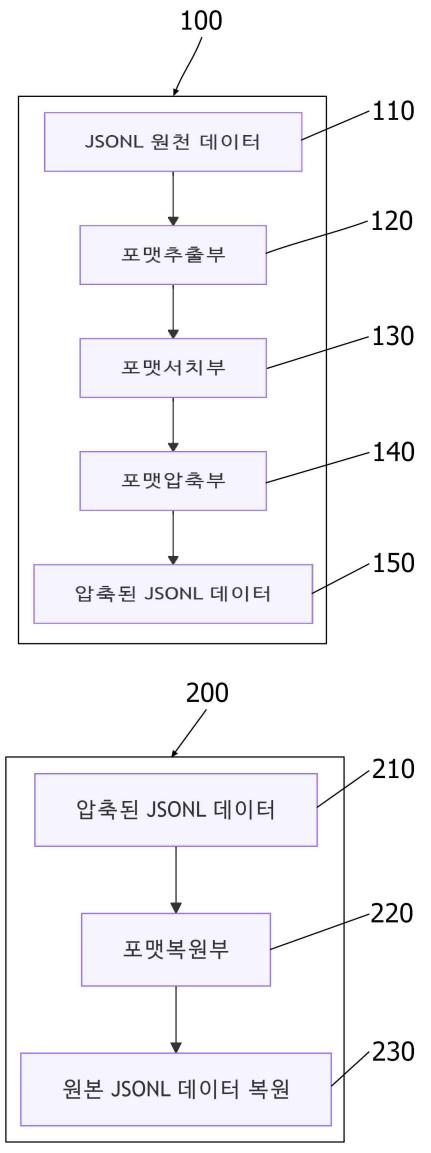
- 필요 시 세그먼트 경계를 정의하고 복수의 사전을 운용할 수 있으며, 이 경우 버전 관리를 적용한다.
 - 긴 스트리밍 환경에서는 SYNC 마커를 삽입하여 오류 격리 및 재동기화를 지원할 수 있다.
 - 해시/체크섬, 서명 등 보안 기능을 선택적으로 부가할 수 있다.

발명의 효과:

- JSONL 포맷 호환성 유지로 가독성과 범용성을 확보한다.
- 사전 기반 압축으로 전송·저장 효율성을 개선한다.

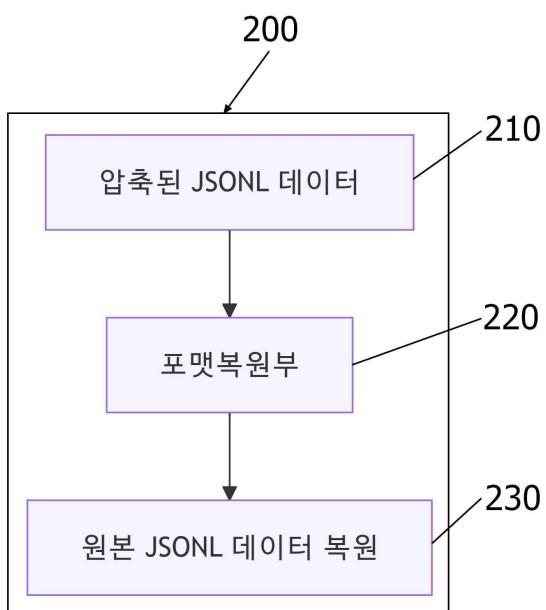
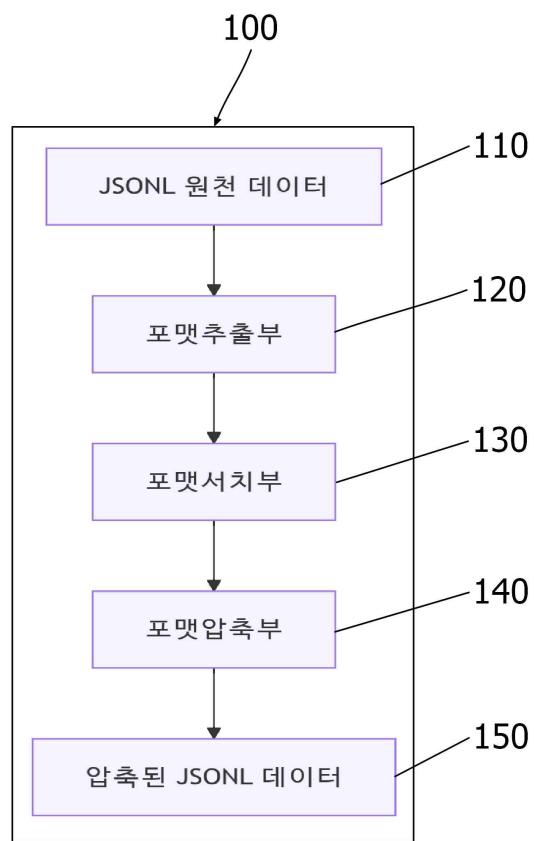
- 라인 독립성 활용으로 부분 복원 · 병렬 복원이 가능하다.
- AI 학습 데이터, 로그, IoT 등 다양한 환경에서 활용할 수 있다.

【대표도】

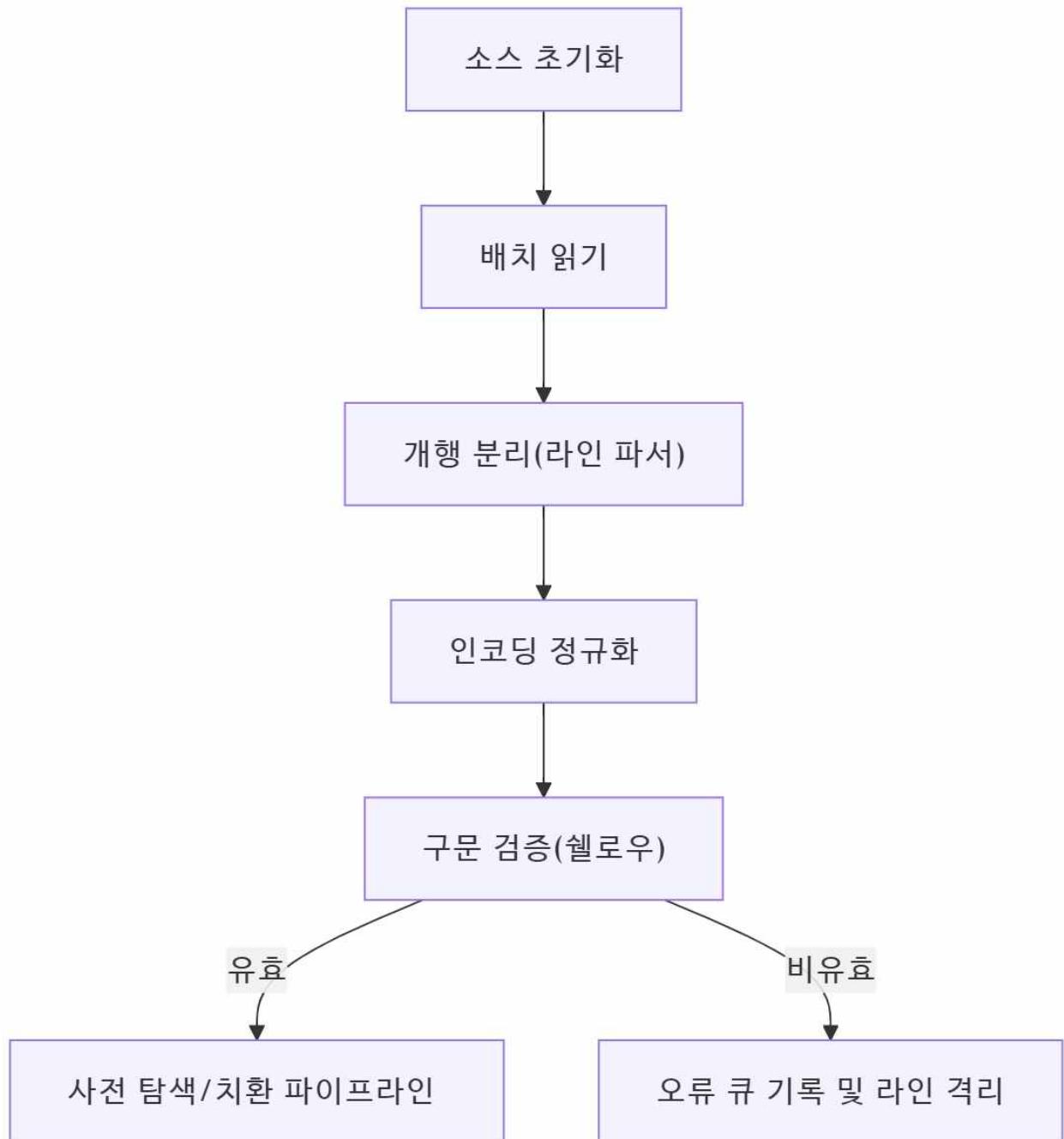


【도면】

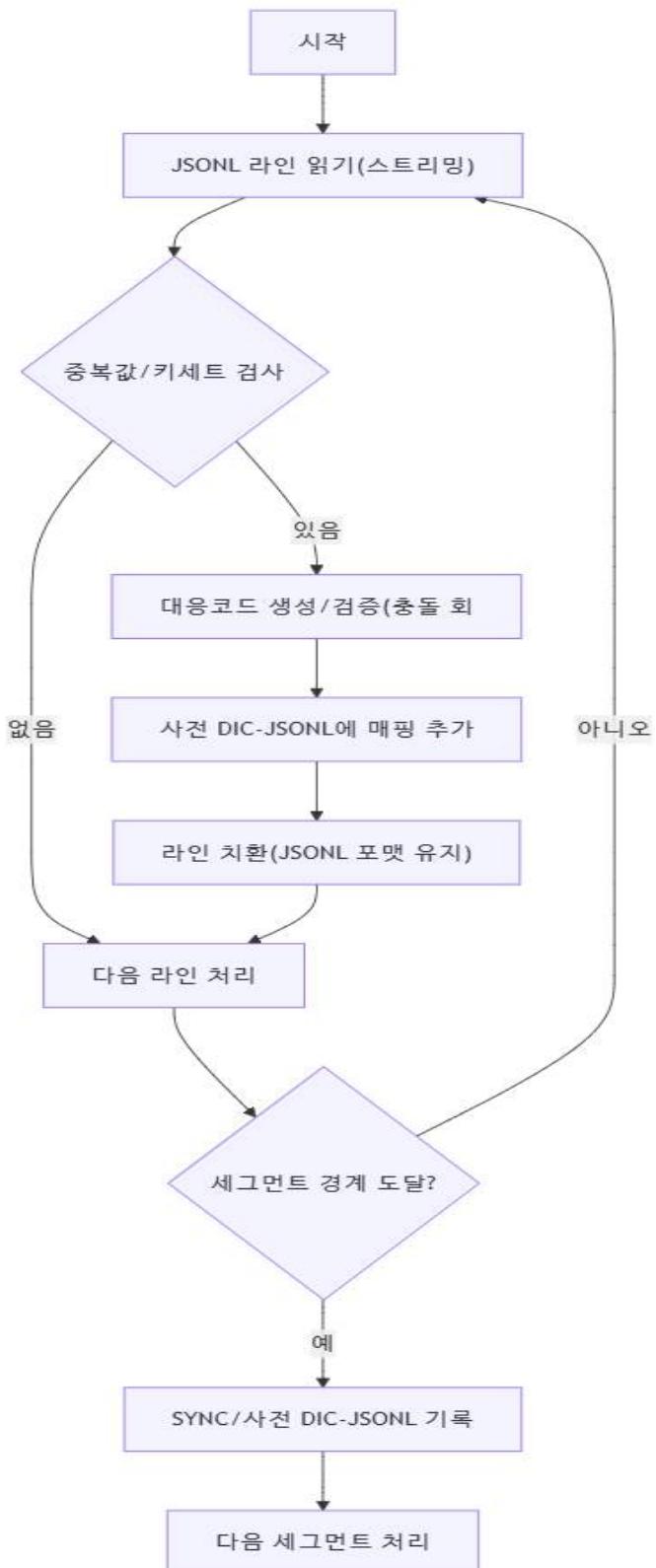
【도 1】



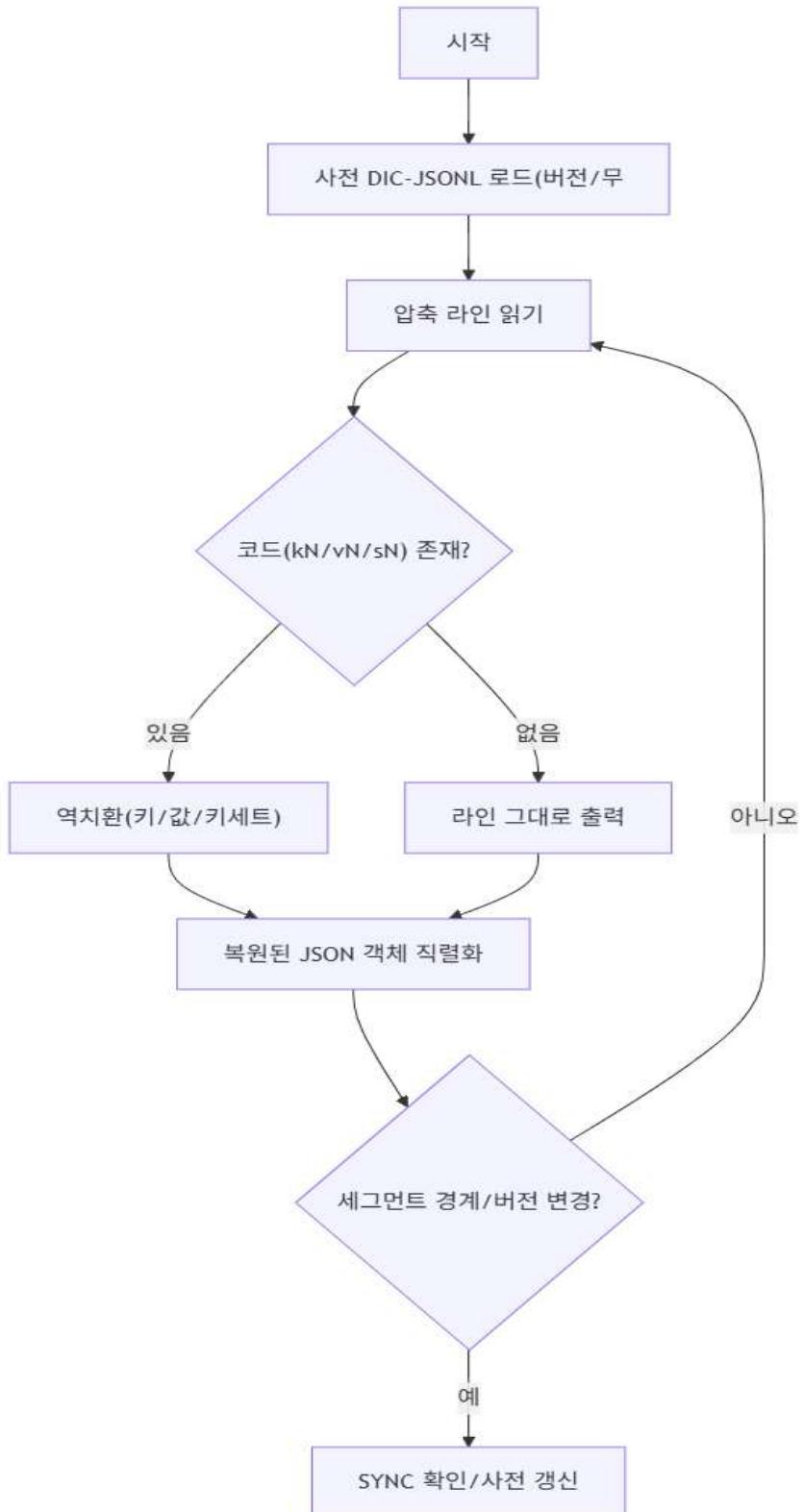
【도 2】



【도 3】



【도 4】



【도 5】

{"also_buy": [], "also_view": [], "asin": "B00ADZ3WUM", "brand": "NSI", "category": ["Automotive", "Exterior Accessories", "Bumper Stickers, Decals & Magnets"], "description": ["Bumpersticker: A day without sunshine is like, well, night. 11\"x2.5\" sticker. Art by NSI"], "feature": ["Official Licensed Die-Cut Sticker Designed by NSI", "Premium Decal is Great for All Surfaces; Skateboards, Windows, Walls, Desks, Doors, Automobiles, Motorcycles, Bumpers, Computers, Laptops, Tablets, Phones, Guitars & Etc.", "Extra Durable, Remarkable Adhesive Long Lasting Vinyl Die-Cut Screen Print Sticker, Can Withstand Any Indoor/Outdoor Weather, Water Protected, UV Coated", "Unique Gift Idea Suitable for Fans & Others", "Individual Protective Packaging."], "image": [], "price": "\$4.68", "title": "NSI - A Day Without Sunshine is Like, Well, Night! - Bumper Sticker", "main_cat": "Automotive"}
{"also_buy": [], "also_view": [], "asin": "B005VII5IU", "brand": "General Motors", "category": ["Automotive", "Replacement Parts", "Body & Trim", "Body", "Valances"], "description": ["This is the official Genuine General Motors Parts replacement part for your vehicle."], "feature": ["This is the official Genuine General Motors Parts replacement part for your vehicle."], "image": [], "price": "\$213.16", "title": "Genuine GM Parts 10341533 Rear Bumper Valance Panel", "main_cat": "Automotive"}

【도 6】

{"_type": "dict_header", "version": 1, "dict": {"k": {"k1": "also_buy", "k2": "also_view", "k3": "asin", "k4": "brand", "k5": "category", "k6": "description", "k7": "feature", "k8": "image", "k9": "price", "k10": "title", "k11": "main_cat"}, "v": {"v1": "Automotive", "v2": "Exterior Accessories", "v3": "Bumper Stickers, Decals & Magnets", "v4": "Replacement Parts", "v5": "Body & Trim", "v6": "Body", "v7": "This is the official Genuine General Motors Parts replacement part for your vehicle.", "v8": "Lights & Lighting Accessories", "v9": "Lighting Conversion Kits", "v10": "Headlight & Tail Light Conversion Kits", "v11": "Lighting Assemblies & Accessories", "v12": "Brake & Tail Light Assemblies", "v13": "Switches", "v14": "Towing Products & Winches", "v15": "Brake System", "v16": "Brake Pads", "v17": "Motorcycle & Powersports", "v18": "Protective Gear", "v19": "Helmets", "v20": "RV Parts & Accessories", "v21": "Covers", "v22": "RV & Trailer Covers", "v23": "Hitch Accessories", "v24": "Filters", "v25": "Transmission & Drive Train", "v26": "Electrical", "v27": "Interior Accessories", "v28": "Starters & Alternators", "v29": "Spectra Premium", "v30": "Fuel System", "v31": "Quality-Built", "v32": "Alternators & Generators", "v33": "Alternators", "v34": "Quality-Built Alternators are remanufactured for a perfect fit to your engine and are redesigned to operate with every turn of the key for reliable performance. Housings are 100% blasted clean, all mounting threads are inspected and re-tapped for easy installation and consistent torque. Terminals are 100% OE-quality; high-temp insulators make connections secure and reliable. Rotors are electronically tested and coated with high dielectric insulation to ensure maximum durability and charging performance. Bearings are 100% inspected or new, with high-temperature grease for reduced heat and friction. Stators are electronically tested for maximum insulation quality and phase balance. Rectifiers are 100% load tested to ensure alternator durability and charging performance. Brushes & Springs are 100% new to ensure long life and charging performance. Pulleys are 100% plated or powder coated for maximum corrosion protection. Slip Rings are precision-polished for maximum output. Voltage Regulators are 100% tested for correct operating voltage.", "v35": "Quality-built Premium alternators are remanufactured to exacting OE specifications or better for maximum performance. They undergo rigorous manufacturing and testing processes to deliver unsurpassed reliability with every turn of the key. Our processes incorporate Durability Testing, Root Cause Analysis and Lean Manufacturing principles; all done in ISO/TS 16949 certified manufacturing facilities. By replacing every component prone to wear with 100% new parts, and end-of-line testing of every unit, Quality-built delivers what a rebuilt unit can't - peace of mind!", "v36": "Quality-Built Premium Alternators undergo a rigorous remanufacturing process to deliver unsurpassed reliability with every turn of the key", "v37": "Our extensive remanufacturing process incorporates Durability Testing, Root Cause Analysis and Lean manufacturing principles", "v38": "By replacing every component prone to wear with 100% new parts, QB delivers what a rebuilt unit can't - peace of mind", "v39": "Comprehensive 1 year Roadside Assistance if unit stops working for any reason", "v40": "Toll-free Quality-Built Technical Support Hotline staffed by ASE-certified master technicians", "v41": "ACDelco", "v42": "Bearings & Seals", "v43": "Bearings", "v44": "Monroe", "v45": "Shocks, Struts & Suspension", "v46": "Shocks & Struts", "v47": "Shocks", "v48": "Motorcycle & Powersports", "v49": "Parts", "v50": "Body & Frame Parts", "v51": "B0002UEN1A", "v52": "Gaskets", "v53": "Air Filters & Accessories", "v54": "https://images-na.ssl-images-amazon.com/images/I/412XCLbr6HL_SS40_.jpg", "v55": "Air Filters", "v56": "\$9.95", "v57": "Custom Fit", "v58": "Spark Plugs & Wires", "v59": "Spark Plugs", "v60": "Sway Bars & Parts", "v61": "Tools & Home Improvement", "v62": "\$9.99", "v63": "Wheel", "v64": "Hub Assemblies", "v65": "Engines & Engine Parts", "v66": "Accessories", "v67": "Performance Parts & Accessories", "v68": "DuraGo", "v69": "Replacing worn hub assemblies with DuraGo Premium Hub Assemblies will reduce friction and improve fuel economy. We use the highest quality components to assemble each hub unit. Each Hub Assembly is individually heat treated utilizing the latest digital machinery. Frequency and current are closely monitored to ensure proper heat treatment. After the induction process, each unit is tempered to improve ductility and toughness and increase impact resistance. Our Hub Assemblies are 100% tested to ensure each unit meets our high standards. All DuraGo Hub Assemblies are proudly manufactured to industry leading quality standards and meet ISO/TS 16949:2002.", "v70": "Forged flange: Provides precise vehicle fit", "v71": "Precise ABS signaling: Ensures proper brake function", "v72": "Multi-lip seals: Provides water and debris contamination protection", "v73": "Superior bearing design: Ensures smooth vibration free operation", "v74": "Induction hardened raceways: Provides vibration free performance and durability", "v75": "Paint, Wall Treatments & Supplies", "v76": "Wall Stickers & Murals", "v77": "Tires & Wheels", "v78": "Accessories & Parts", "v79": "Trailer Accessories", "v80": "Hardware", "v81": "Connectors", "v82": "Evan Fischer", "v83": "Fits the following vehicles:", "v84": "FREE 1-year UNLIMITED mileage warranty coverage on Evan-Fischer items purchased thru AUTO PARTS GIANT Store", "v85": "Problem Solver bushings for superior performance", "v86": "\$8.18", "v87": "Tools & Equipment", "v88": "Dorman"}, "s": {"s1": ["also_buy", "also_view", "asin", "brand", "category", "description", "feature", "image", "price", "title", "main_cat"]}], "entryCount": {"keys": 11, "values": 88}, "keysets": 1}, "hash": "1e62614b13e544de"}

{"s1": [], "v1": "B00ADZ3WUM", "NSI": ["v1", "v2", "v3"], "v": ["Bumpersticker: A day without sunshine is like, well, night. 11\"x2.5\" sticker. Art by NSI"], "v": ["Official Licensed Die-Cut Sticker Designed by NSI", "Premium Decal is Great for All Surfaces; Skateboards, Windows, Walls, Desks, Doors, Automobiles, Motorcycles, Bumpers, Computers, Laptops, Tablets, Phones, Guitars & Etc.", "Extra Durable, Remarkable Adhesive Long Lasting Vinyl Die-Cut Screen Print Sticker, Can Withstand Any Indoor/Outdoor Weather, Water Protected, UV Coated", "Unique Gift Idea Suitable for Fans & Others", "Individual Protective Packaging."], "v": "\$4.68", "NSI - A Day Without Sunshine is Like, Well, Night! - Bumper Sticker", "v": []}, {"s1": [], "v1": "B005VII5IU", "General Motors": ["v1", "v4", "v5", "v6", "Valances"], "v": ["General Motors", "Valances"], "v": "\$213.16", "v": "Genuine GM Parts 10341533 Rear Bumper Valance Panel", "v": []}]