**웹 스크래핑을 활용한 사회문제에 대한 언론 매체의 관심도 분석:**

**미세먼지 문제를 중심으로**

**김은배**

본문 요약

2013년 세계보건기구에서 ‘미세먼지’를 1급 발암물질로 규정했고, 국내에서는 2015년부터 전국적인 초미세먼지 관측이 시작됐다. 이로 인해, 미세먼지를 향한 언론의 관심이 증폭되었고, 곧 일반 대중의 인식 변화를 가져왔다.

이에 본 논문에서는 실제 미세먼지 농도와 이와 관련된 언론의 보도 지표를 웹 스크래핑 기술을 활용하여 추출할 예정이다. 그 후, 데이터 전처리와 시각화를 통해 두 측정값 사이의 상관관계를 확인할 것이다. 만약, 이 연구에서 의미 있는 분석 결과가 도출된다면, 언론 매체의 미세먼지 관련 보도의 방향성과 그 의도를 발견할 수 있는 계기가 될 것이다.

1. 서론

미세먼지는 직경에 따라 PM10, PM2.5로 구분되며, PM10은 1000분의 10mm, PM2.5는 1000분의 2.5mm보다 작은 크기의 입자로 이루어져 있다. 미세먼지의 성분은 발생 지역, 기상 조건, 계절 등에 따라 달라질 수 있으나, 가장 문제가 되는 것은 매연, 배기가스, 날림먼지 등의 인위적 미세먼지이다. (환경부와 친해지구, 2018) 특히, 초미세먼지라 불리는 PM2.5의 경우 입자 크기가 작아 폐포 깊숙한 곳까지 먼지가 침투하기 때문에 각종 호흡기계 질환 및 폐 기능 저하를 초래할 수 있다. (국가지표체계, 2023)

2013년 세계보건기구에서는 미세먼지를 1급 발암물질로 규정했으며, 우리나라의 경우 PM2.5에 대한 대기 환경 기준을 2011년에 신설하여, 2015년부터 전국 규모 측정값을 공개하고 있다. 이처럼, 건강과 직결되는 미세먼지 농도 발표는 언론의 대대적 관심을 받기 시작했고, 대중들 또한 미세먼지에 대한 경각심을 가지게 되었다.

여기서 우리가 주목해야 할 부분은 미세먼지와 같은 위험보도가 대중에게 끼치는 영향력이다. 언론은 파급력 있는 사건을 보다 자극적으로 보도하기 위해 위험을 강조하는 특성이 있다. 결국 같은 사건이라도 보도 방식에 따라 대중의 가치판단, 행동양식에 각기 다른 영향을 끼칠 수 있게 된다. 또한, 편향적이며 선정적인 보도는 공포감 및 부정적 감정을 유발하여 사회적 혼란을 야기할 가능성이 있기에 주의해야 한다. (김영욱, 위험 불통 사회, 2021)

따라서 본 연구에서는 실제 미세먼지 농도의 변화와 언론의 관심도 간의 상관관계를 웹 스크래핑을 통해 살펴볼 예정이다. 만약 실제 미세먼지 수치와 언론의 관심도가 비슷한 추세라면 언론은 미세먼지와 같은 재난 · 재해에 올바르게 반응(대응)했다고 볼 수 있다. 반대로, 두 수치 사이에 큰 연관성이 존재하지 않을 시 언론은 미세먼지 보도 방식을 재고해야 하며, 대중들 또한 이에 대한 시정을 촉구해야 한다.

2. 연구방법

1) 자료 추출 방법

본 연구에서는 미세먼지 관련 자료를 웹 스크래핑 자동화 도구인 ‘Selenium’을 사용하여 수집했다. 여기서 웹 스크래핑(Web Scraping)이란, 웹사이트에서 정보를 추출하고, 정돈된 컴퓨터 문서로 만드는 작업이다. (Cambridge dictionary, “web scraping”)

추출한 데이터는 Python을 기반으로 한 라이브러리인 Pandas와 Plotly를 사용하여 테이블 및 그래프로 시각화했다. Pandas는 대용량 데이터를 처리할 때 매우 유용한 라이브러리이며, Plotly는 상호작용이 가능한 40개 이상의 차트를 제공하는 오픈소스 그래픽 라이브러리이다.

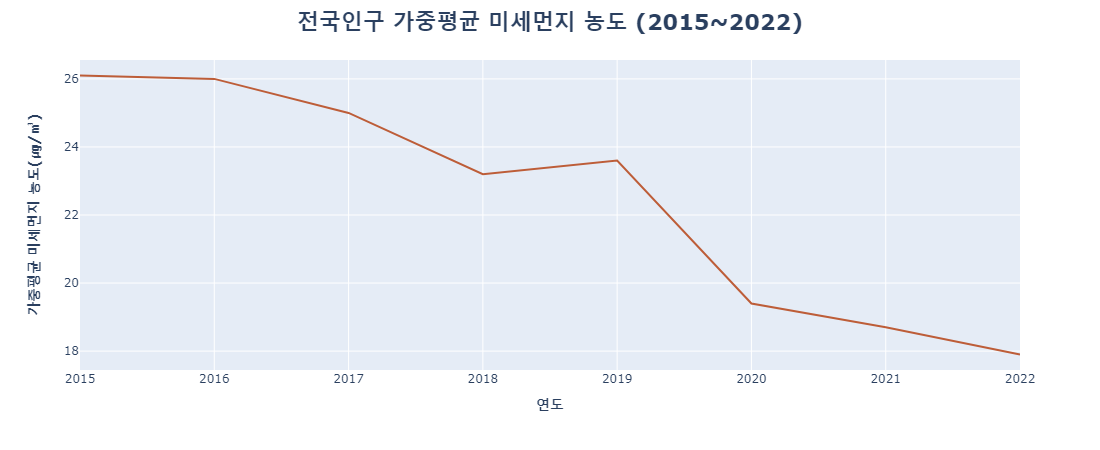
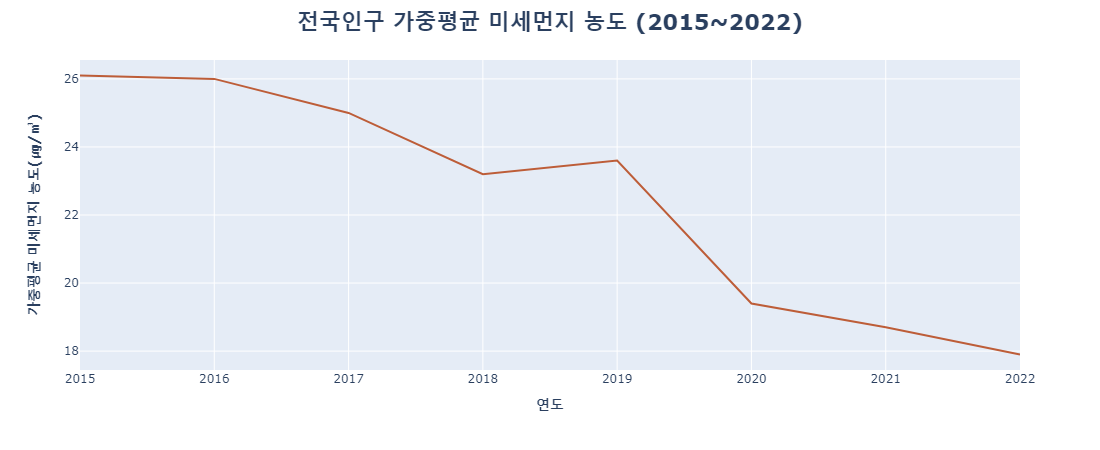
2) 수집한 데이터

위의 도구들로 먼저 국가지표체계의 ‘2015~2022년 전국인구 가중평균 미세먼지 농도’를 수집했다. 17개 시·도의 연평균 미세먼지(PM2.5)농도 인구가중 평균값을 선 그래프로 만들어 연도별 미세먼지 농도 변화를 확인할 수 있도록 했다.

그 다음으로 추출한 값은 ‘월별 미세먼지 기사 개수’이다. 구글의 뉴스페이지로 이동하여 키워드를 “미세먼지”로 설정한 후 검색하여 나온 결과이며, 2015년 1월부터 2023년 11월까지의 월별 기사 개수를 수집했다. 이 데이터들은 테이블로 변환하거나, 연도별로 묶어서 막대 그래프나 선 그래프로 제작해 연도별 증감율을 확인할 수 있도록 했다.

그 밖에도, 국가 통계 포털(KOSIS)에서 미세먼지(PM2.5) 월별 대기오염도를 2015년 1월~2023년 2월까지 CSV 파일로 다운로드했으며, 이를 테이블로 변환하였다.

3) 그래프 분석



**<그림 1. 2015년~2022년 전국인구 가중평균 미세먼지 농도 그래프>**

<그림 1> 그래프는 2015년부터 2022년까지 측정된 전국인구 가중평균 미세먼지 농도를 나타내고 있다. 전국적인 관측이 시작된 2015년 이래로, 미세먼지 농도는 지속적인 감소 추세를 기록하고 있다. 물론 2019년에는 전년도에 비해 미세먼지 농도가 소폭 상승했으나, 2020년에 다시 획기적으로 그 수치가 개선되었다.

정부는 2020년의 미세먼지 농도 개선폭을 ‘정부정책으로 인한 국내 배출량 감축 효과’, ‘코로나 19로 말미암은 사회 · 경제적 활동의 감소’, ‘양호한 기상 상황’, ‘중국의 미세먼지 개선 추세’ 등으로 그 이유를 추정하고 있다. (대한민국 정부,” 올해 미세먼지가 획기적으로 개선된 것은 코로나19의 영향 아닌지?”, 2020) 그 이후로도 미세먼지 농도는 계속해서 감소했으며, 2022년에는 관측 이래 가장 낮은 수치인 17.9㎍/㎥를 기록했다.

**텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

자동 생성된 설명<그림 2. 2015~2023 연도별 미세먼지 기사 막대 그래프 (2023년은 12월은 수치에서 제외)>**

**텍스트, 도표, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명**

**<그림 3. 연도별 미세먼지 기사의 증가율 (2023년은 12월은 수치에서 제외)>**

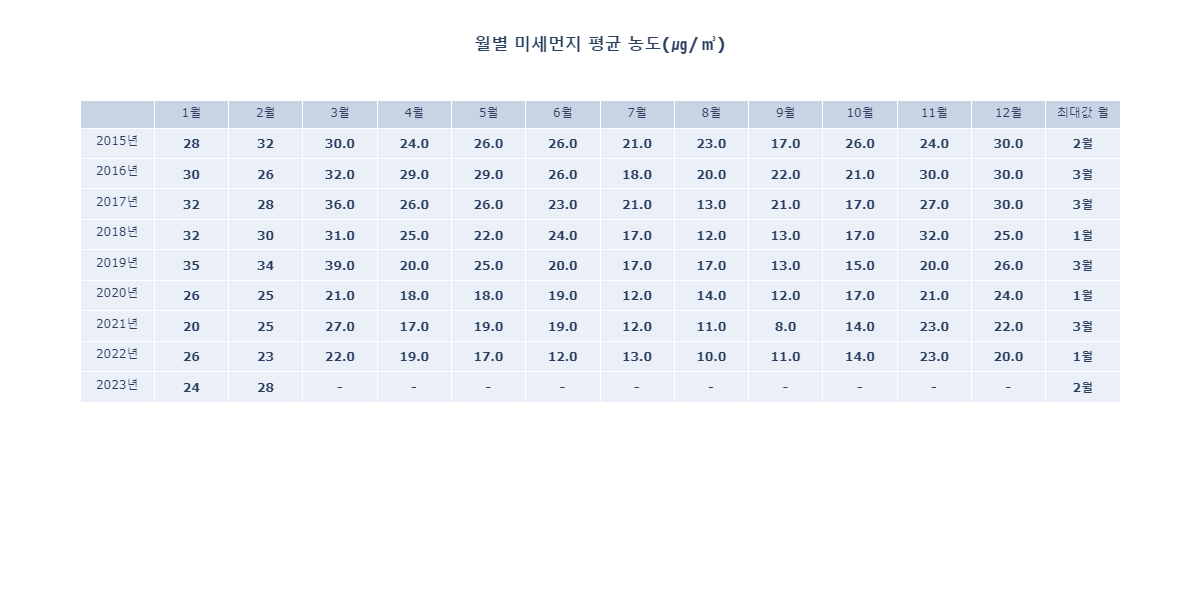
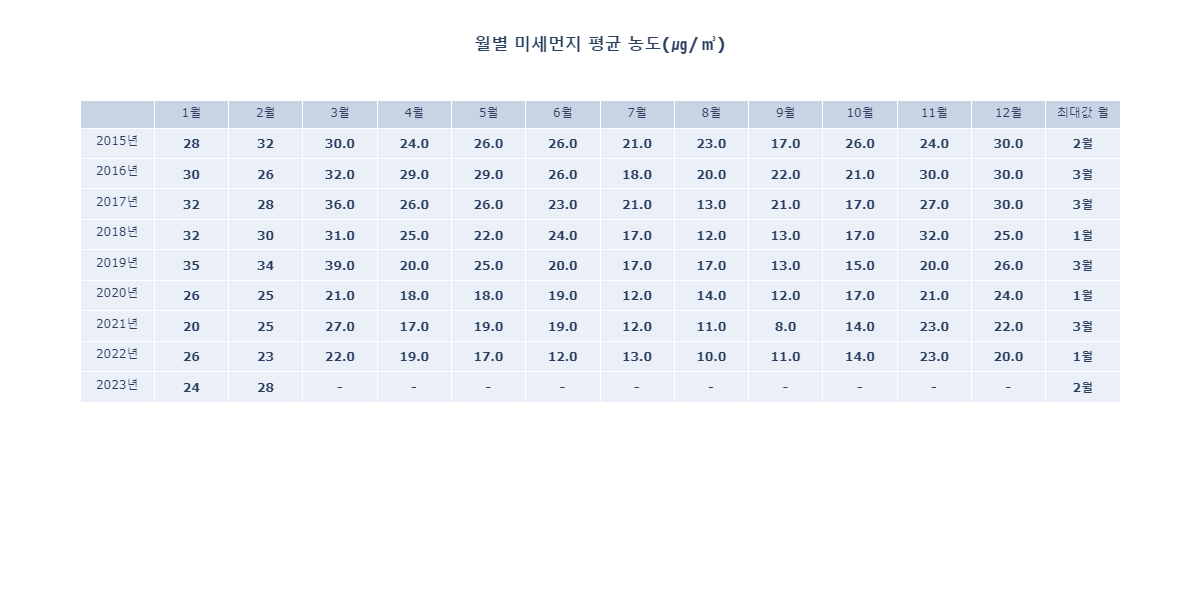
<그림 2> 그래프는 2015년부터 2023년(12월은 제외)까지의 연도별 미세먼지 기사의 수를 막대 형태로 보여주고 있다. 매해 미세먼지 관련 기사 수가 증가하는 것을 볼 때, 언론 매체는 여전히 미세먼지를 대중에게 호소 가능한 주제로 인식함을 알 수 있다. <그림 3> 그래프는 선 그래프로 전년 대비 미세먼지 기사 수의 증감을 보여주고 있다.

여기서 주목해야 할 지점은 2019년, 2022년 그리고 2023년 미세먼지 기사 수의 변화율이다. 먼저 2019년에는 2018년 대비 64%의 비율로 기사 수가 증가했고, 그 원인은 <그림 1> 그래프를 통해 추정 가능하다. 이전까지 감소 추세에 있던 미세먼지 농도가 2019년 상승했고, 이에 위험을 감지한 언론 매체가 미세먼지 보도에 집중한 것으로 보인다.

2022년은 15%의 증가율을 기록한 해로, 6% 대를 기록한 2021년에 비해 그 비율이 상승했다. 분명 미세먼지 농도가 관측 이래 최저치를 기록한 해임에도 향후 예상되는 부정적 전망이 보도량에 영향을 끼친 것으로 보인다. (조선일보, “中 석탄 발전량 급증… 올겨울 초미세먼지 심해질 듯”, 2022.12)

마지막으로 2023년은 12월을 제외하고 측정했음에도 이미 전년 대비 75%의 증가율을 기록했다. 이처럼 언론 매체가 일제히 우려를 표명하는 이유는 2022년보다 심화된 대기오염 관련 전망이 있었기 때문이다.

국내의 경우, 정부가 코로나 19의 엔데믹을 발표(BBC뉴스,” 한국, 코로나19 ‘엔데믹’ 선언…무엇이 달라지나?”, 2023)했고, 중국 또한 제로 코로나 종결(BBC뉴스, “코로나19: 중국, 1월 8일부터 '해외 입국자 시설 격리' 폐지”, 2023)을 선언하며 국내 · 외 각종 산업이 재개될 조짐을 보이고 있었다. 이에 더해, 2023년 겨울에는 엘니뇨와 같이 대기 환경을 악화시키는 기상 여건(BBS뉴스, “한 총리, 올 겨울 미세먼지 농도 높아질 것으로 우려”, 2023)이 예상되면서 언론 매체가 미세먼지 관련 이슈에 더욱 민감하게 반응하고 있다.



**<그림 4. 연도별 · 월별 미세먼지 평균 농도 테이블(2023년 2월까지)>**

<그림 4> 테이블은 2015년 1월부터 2023년 2월까지 월별 미세먼지 평균 농도를 나타내고 있다. 마지막 열에는 각 연도 별로 가장 높은 미세먼지 농도를 기록한 달을 기입해 두었다.

미세먼지 농도는 보통 중국발 미세먼지와 대기 안정을 이유로 겨울철에 그 수치가 가장 높은 것으로 알려져 있는데, (헬스조선, ”겨울철에 더 심해지는 미세먼지 대처법”, 2016) <그림 4> 테이블 또한 1월에서 3월 사이에 가장 높은 미세먼지 농도를 기록하고 있다.



**<그림 5. 월별 미세먼지 기사 개수 테이블(2023년 12월 제외)>**

한편, 월별 미세먼지 기사 개수를 기록한 <그림 5> 테이블도 마지막 열에 연도별 최대값을 기입했다. 그 결과, 2023년을 제외하고 12월이 4개의 연도(2015년, 2016년, 2018년, 2020년)에서 최대값을 기록했으며, 나머지 4개의 연도에서는 각기 다른 달에 최대값을 기록했다.

월별 미세먼지 기사의 추이를 사전에 예상해 봤을 때, 미세먼지 농도가 최대인 1~3월에 언론 매체가 가장 많은 보도를 할 것이라 추측해볼 수 있다. 하지만 실제로 <그림 4>와 <그림 5> 테이블의 측정값을 확인했을 때, 유의미한 상관관계를 발견할 수 없었다. 이는 언론 매체가 미세먼지 관련 보도를 결정할 때, 실제 추세에 따라 보도의 방향을 설정하지 않으며, 이해관계에 따라 해당 담론을 설정하거나 배제한다는 것으로 해석할 수 있다.

3. 결론

위의 연구를 통해, 실제 미세먼지 농도 변화(<그림 1>, <그림 2> 2019년 수치)뿐만 아니라, 미세먼지와 관련된 부정적 전망(2022 ~ 2023년)이 언론의 관심도에 영향을 미친다는 것이 증명됐다. 이에 더해, 월별 미세먼지 농도와 월별 미세먼지 기사 수가 큰 상관관계를 보이지 않았다는 사실은, 언론 매체가 이해관계에 따라 의도적으로 미세먼지 관련 이슈를 선택하고 배제할 수 있다는 방증이 된다.

여론을 형성하는 대중은 이러한 사실을 항상 인지하고 있어야 하는데, 언론 매체가 어떤 의제에 집중하느냐에 따라 대중들의 관점 또한 달라지기 때문이다. 이는 “언론이 주목하지 않는 의제는 대표 의제로 부상하기 어렵다.” 라는 말(McCombs & Shaw, 1972; Rogers & Dearing, 1988)과도 일맥상통한다. 더 나아가, 언론 매체는 프레이밍(framing)이라는 과정을 통해 특정 시각을 선택하고 강조하며 개인의 도덕적 가치판단에도 영향(McCombs, 2004)을 주게 된다. 결국, 잠재적인 위험 상황을 빌미로 선정적인 보도를 하는 방식(Kitzinger, 1999)은 미세먼지에 대한 대중의 불안을 야기한다.

언론 매체가 대중에게 불안감을 조성하는 이유는 바로 ‘프로모션 기사’의 존재 때문이다. 최근 미세먼지 관련 기사에서는 특정 상품과 서비스가 언급되는 사례가 점차 늘어나고 있다. 이는 미세먼지에 대한 원인 규명과 해결방안 제시라는 언론 본연의 임무보다, 공포감 조성을 통한 비즈니스에 치중한 결과이다. 이렇게 왜곡된 정보 전달 방식은 결국 언론에 대한 전반적인 신뢰도 하락의 원인이 되고 있다. (김영욱 외 3명, 2014)

위와 같은 문제를 해결하기 위해서는 언론 매체 스스로 쇄신을 촉구해야 한다. 물론 광고 · 홍보형 기사의 경우, 단기적 관점에서 언론의 이익을 증진시킬 여지가 있다. 하지만 전반적인 신뢰도 하락은 결국 대중의 외면을 받게 되며, 언론 매체의 주 수익원인 CPV(Cost per view), CPM(Cost per Mille) 산정 시에 장기적인 손해를 가져올 것이다. 따라서 언론 매체는 대중과 공익의 관점에서 공정하고 독립적인 보도를 추구해야 한다. (한국언론진흥재단, “언론윤리헌장”)

참고문헌

◾ 환경부와 친해지구, 미세먼지, 너는 누구니? 미세먼지의 모든 것, 2018.6.19. 11:11, <https://blog.naver.com/mesns/221302000466>

◾ 미세먼지 농도, 국가지표체계, 2023-04-04, [https://www.index.go.kr/unify/idx-info.do?idxCd=8058](https://www.index.go.kr/unify/idx-info.do?idxCd=8058%20)

◾ 환경부, “올해 미세먼지가 획기적으로 개선된 것은 코로나19의 영향이 아닌지?”, 2020-11-27, <https://www.airkorea.or.kr/portal/web/board/12/1652/?pMENU_NO=144&page=2>

◾ 김영욱 (2021). 위험 불통 사회. 서울: 이화여대출판원

◾ Cambridge dictionary, web scraping, 2023.12.04, https://dictionary.cambridge.org/ko/%EC%82%AC%EC%A0%84/%EC%98%81%EC%96%B4/web-scraping

◾ 박상현, 장근욱, “中 석탄 발전량 급증… 올겨울 초미세먼지 심해질 듯”, 2022.11.11 05:50, <https://www.chosun.com/national/transport-environment/2022/11/11/4MYN7NMNVBADHHCFA4BMY5WVCY/>

◾ BBC 뉴스, “한국, 코로나19 ‘엔데믹’ 선언…무엇이 달라지나?”, 2023.05.11, <https://www.bbc.com/korean/articles/c2q1gpelyz3o>

◾ 프란시스 마오, 판 왕, “코로나19: 중국, 1월 8일부터 '해외 입국자 시설 격리' 폐지”, 2022.12.27, <https://www.bbc.com/korean/news-64099927>

◾ 박원식, “한 총리, 올 겨울 미세먼지 농도 높아질 것으로 우려”, 2023.11.24 10:43, <https://news.bbsi.co.kr/news/articleView.html?idxno=3133689>

◾ 황인태, “겨울철에 더 심해지는 미세먼지 대처법”, 2016.12.05 16:28, <https://m.health.chosun.com/svc/news_view.html?contid=2016120501164>

◾ 김영욱 외 3명, 언론은 미세먼지 위험을 어떻게 구성하는가?: 미세먼지 위험보도 프레임과 정보원 분석, p.5, 2014.12.14

◾ McCombs, M. E. (2004). Setting the agenda—The mass media and public opinion. Malden, MA: Blackwell

◾ McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. Public Opinion Quarterly, 36(2), 176-187.

◾ Rogers, E. M., & Dearing, J. W. (1988). Agenda-setting research: Where has it been, where is it going. In J. Anderson(Ed.), Communication Yearbook 11 (pp.555-594). Beverly Hills, CA: Sage

◾ Kitzinger, J. (1999). Researching risk and the media. Health, Risk & Society, 1(1), 55-69.

◾ 언론윤리헌장,한국언론진흥재단, 2021.01.26, <https://www.kpf.or.kr/front/board/boardContentsView.do?board_id=291&contents_id=610b0895693a446ab7e5b09ddd859c95&miv_pageNo=&miv_pageSize=15&total_cnt=&LISTOP=miv_pageSize%3D15%3B_SERIALISVALID%3DT%3Bmiv_pageNo%3D1%3B&mode=W&p_reg_userno=&searchkey=A&sear>