

OPINION

연구위원
노성호AI의 시대에 기업 재무 공시의 구조화가
가지는 의미*

상장기업과 투자자 간의 소통 경로 중 가장 기초가 되는 재무 공시는 제도적인 검증 절차를 거쳐 기업의 재무 상태에 대한 포괄적이고 신뢰성 있는 정보를 제공한다는 점에서 그 중요성은 크다. 그러나 공시자료의 양적 성장과 복잡화 추세에 비추어 보았을 때 투자자가 방대한 양의 정보를 효율적으로 처리하여 의사 결정에 반영하기는 매우 어려운 과제이다. 이에 대응하여 자연어 처리 모형 등 AI 기술을 활용한 공시자료의 수집 및 처리의 자동화 경향은 지난 수 년간 빠르게 진행되고 있다.

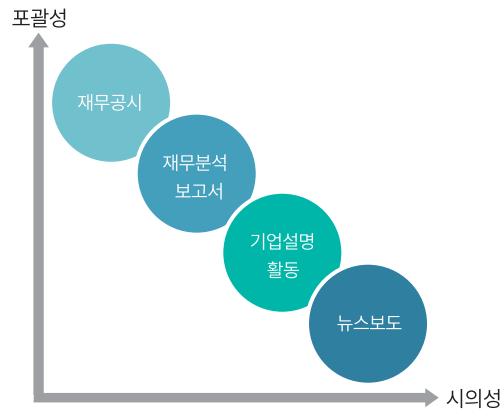
재무 공시의 구조화 및 디지털화 노력은 이와 같은 '기계' 독자의 증가와 무관하지 않다. XBRL 도입은 복잡한 재무 계정과 주석을 일관된 체계로 분류하여 기계 가독성을 높이고 있으며 더불어 선진국 자본시장을 중심으로 디지털 공시자료에 대한 접근성을 강화할 수 있는 아카이브 등의 인프라 구축 노력도 지속적으로 이루어지고 있다. 이와 같은 노력은 궁극적으로 재무 공시의 구조화가 정착하고 디지털 접근성을 제고하여 재무 정보를 투자자가 효율적으로 소비하여 시의성 있게 활용하도록 하는데 기여할 것으로 기대한다. 이를 실현하기 위하여 기계 가독성(machine readability)에 대한 이해를 바탕으로 재무 공시에 대한 온라인 접근성을 강화하는 것과 더불어 재무 공시의 유의성을 높이는 노력이 지속되어야 한다. 더불어 디지털 구조화 공시 도입 초기에 집중되는 비용과 비대칭적인 정보의 확산을 최소화하기 위하여 투자자와 기업의 적극적인 소통과 정책적 지원이 필요할 것이다.

기업 재무 공시의 복잡화

상장기업이 투자자와 만나는 다양한 방식 중에서 가장 기본이 되는 것은 재무 공시(financial disclosure)일 것이다. <그림 1>은 투자자가 상장기업에 대한 정보를 얻는 네 가지 대표적인 경로인 재무 공시, 기업설명활동, 재무분석 보고서, 뉴스 보도를 포괄성과 시의성의 관점에서 비교하고 있다. 다른 경로들과 비교하였을 때 재무 공시는 기업의 현황에 대한 정보를 포괄적으로 조망하며 공시제도에 근거한 절차에 따라 검증된 자료라는 점에서 중요도가 매우 높다. 그러나 필연적으로 정보의 양이 방대하고 검증에 시간이 소요되는 만큼 정보의 생산에서 소비까지 오랜 시간이 걸리게 되어 시의성 면에서 다소 약점을 가질 수 있다.

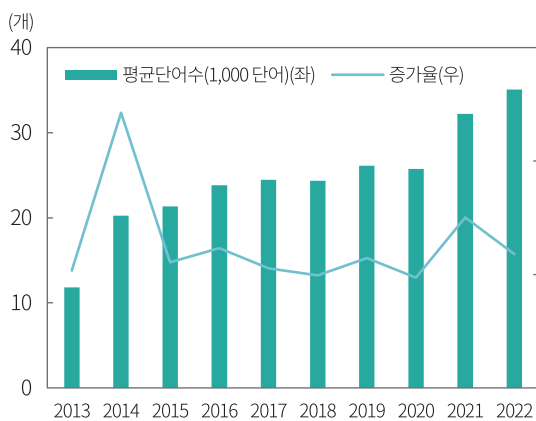
* 본고의 견해와 주장은 필자 개인의 것이며, 자본시장연구원의 공식적인 견해가 아님을 밝힙니다.

〈그림 1〉 기업 정보 공시의 종류



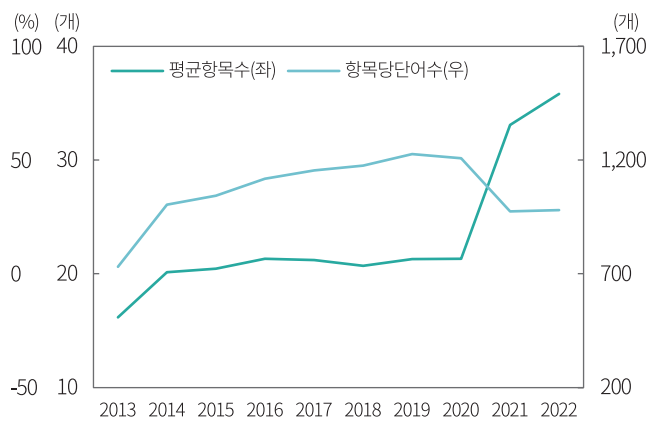
기업 운영 방식과 재무 구조의 복잡화에 따라 공시 정보의 밀도가 높아지고 관련 규제가 세밀하게 발전할수록 공시자료를 독해하는 어려움 또한 가증될 것이다. 회계연도 기준 2013년부터 2022년 사이 금융감독원 전자공시시스템(Data Analysis, Retrieval and Transfer system: DART)에 업로드된 모든 상장기업의 연간 사업보고서를 분석한 결과 문서당 평균 단어 수는 2013년 약 12,000단어에서 2022년 약 35,000단어로 연평균 11% 증가하였다(〈그림 2〉 참조). 더불어 같은 기간 동안 사업보고서의 세부 항목 개수는 평균 16개에서 35개로 2배 이상 증가하였으며 항목당 분량도 약 730단어에서 980단어로 증가하였다(〈그림 3〉 참조). 이는 지난 10여 년간 기업의 재무 공시가 양적으로만 성장한 것이 아니라 다층적인 정보를 높은 밀도로 제공하는 형태로 발전하였음을 보여주고 있다.

〈그림 2〉 사업보고서 분량



자료: 금융감독원 전자공시시스템 및 저자 계산

〈그림 3〉 사업보고서 항목 수



재무 공시의 구조화 및 디지털화 노력

재무 공시의 양적 성장과 복잡화에 대응하여 AI 등을 활용하여 정보 처리의 효율성을 제고하려는 노력이 활발하게 이루어지고 있다. 자연어 처리 모형의 대두와 Loughran and McDonald(2011)의 금융 및 재무 용어 사전 연구 등의 성과가 기반이 되어 EDGAR¹⁾에 대한 기계적인 접근(예를 들어 웹 스크래핑)이 2000년대 초반 이후 급증하였다(Cao et al., 2023). 이는 실무 차원에서 기업이 실적 발표에서 기계 독자까지 고려하여 단어를 선택하고(Wiggleworth, 2020), 재무분석사(financial analyst)가 구조화된 재무 공시의 항목별 태그(tag)를 활용해 예측 정확도를 높이고(Felo et al., 2018), 감독 기관이 업무에 필요한 대량의 정보를 알고리즘을 통해 효과적으로 수집하는(SEC, 2023) 변화로 나타나고 있다.

재무 공시의 구조화 및 디지털 접근성 강화는 이를 통해 기계적인 방식으로 대량의 공시자료를 분석하는 작업의 효율성과 신뢰성을 높일 수 있다는 점에서 중요성을 더하고 있다. <표 1>은 주요국의 공시자료 구조화 및 디지털 접근성에 대한 현황을 요약한 것이다. XBRL의 도입은 유럽을 제외하고 2000년대에 이미 이루어져 재무 공시의 국제적 정합성을 높이려는 노력이 일찍이 진행되었음을 알 수 있다. 반면, 아카이브 및 API(Application Program Interface) 구축을 통하여 공시자료에 대한 디지털 접근성을 강화하는 것은 여전히 많은 나라에서 과제로 남아있다고 볼 수 있다. 국내의 경우 금융감독원에서 2023년 3월 상장기업의 재무제표에 대한 XBRL의 단계적 도입 계획을 발표하였다.²⁾ 이는 공시자료의 디지털 구조화를 통해 국제적 정합성을 제고한다는 측면에서 긍정적인 변화일 것이다.³⁾ 더불어 금융감독원의 DART는 공시자료의 체계적인 검색과 더불어 API를 활용하여 자동화된 자료 검색 및 수집 기능을 제공한다는 점에서 디지털 접근성 또한 해외 주요국 대비 높은 수준이라고 볼 수 있다.

1) Electronic Data Gathering, Analysis, and Retrieval system의 약자로 미국 증권거래위원회(U.S. Securities and Exchange Commission)에서 상장기업에 대한 공시자료의 수집, 검증, 분류 및 배포를 자동화하기 위하여 구축한 디지털 아카이브이다.

2) 금융감독원(2023)

3) 기존 XBRL 관련 규정은 비금융업 상장법인, 그리고 재무제표 본문에만 제한적으로 적용되어 효율성에 대한 의문이 제기되었다(홍지연, 2023).

〈표 1〉 국가별 공시자료 구조화 및 디지털화 현황

구분	세부	미국	유럽	일본	한국
구조화 공시	XBRL 도입 시기	2009년	2020년 ¹⁾	2008년	2010년 ²⁾
	XBRL 적용 범위	재무제표 본문과 주석 일부	재무제표 본문과 주석 일부	재무제표 본문과 주석 일부	재무제표 본문과 주석 일부
	태그(tag) 작성 방식	세부항목 단위 (Detailed Tagging)	목차 단위 (Block Tagging)	세부항목 단위 (Detailed Tagging)	세무항목과 목차 단위 혼합
디지털 접근성	디지털 아카이브	EDGAR		EDINET	DART
	API 도입 여부	사용가능 (2021. 9.)		사용가능	사용가능 (2013. 8.)
	API 사용 가능 항목	기업현황 정기보고서 계정항목		정기보고서	기업현황 정기보고서 재무제표 등

주 : 1) 재무제표 주석에 한하여 2022년 시행으로 도입 시기를 2년 유예하였다.

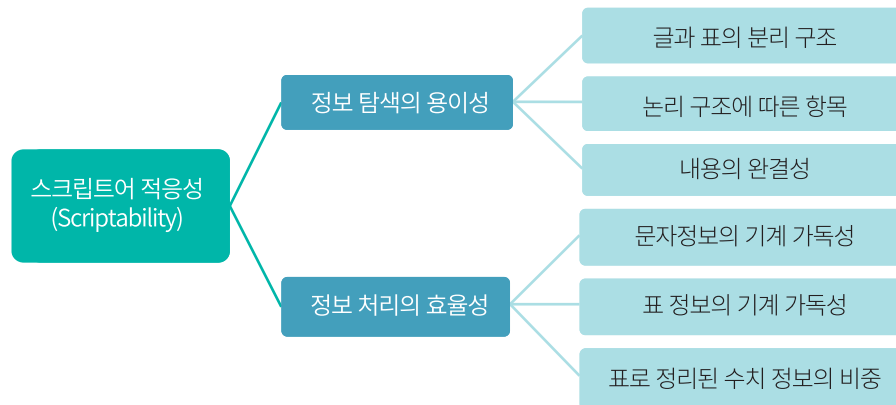
2) 2007년 K-GAAP기반 XBRL의 도입 이후 2010년 IFRS기반 공시시스템을 구축하였으며 2023년부터 금융업 상장법인, 재무제표 주석으로 XBRL 적용 범위를 확대할 예정이다.

자료: 금융감독원(2023), 홍지연(2023), EDGAR, EDINET, DART 홈페이지를 참고하여 작성하였다.

공시자료의 가독성: 인간의 관점과 기계의 관점

공시자료의 구조화 정도가 높아질수록 상대적으로 인간 독자보다 기계의 독해력만 높아질 것인가에 대한 의문이 제기될 수 있다. 이에 대한 답변에 앞서 공시자료의 가독성(readability)에 대하여 기계의 관점에 대한 정의가 필요할 것이다. Allee et al.(2018)은 기계의 입장에서 공시자료의 가독성을 정의하는 지표로서 스크립트어 적응성(scriptability)을 제안하였다(그림 4) 참조).

〈그림 4〉 스크립터 적응성의 개념



자료: Allee et al.(2018)을 바탕으로 재구성

스크립터 적응성은 크게 정보 탐색의 용이성과 정보 처리의 효율성을 포괄하는 개념으로서 제시되었다. 밀도가 높은 문서에서 알고리즘을 통하여 빠르게 원하는 정보를 찾을 수 있다면(정보 탐색의 용이성), 해당 공시자료의 유용성은 그렇지 않은 문서보다 높을 것이다. 정보의 탐색을 쉽게 하기 위한 세부 과제로는 (1) 항목 사이의 논리적 구조와 제목의 연관성을 강화하고, (2) 서술형 정보와 표를 문서에서 쉽게 분리할 수 있도록 하며, (3) 외부 링크 또는 첨부 문서에 의존하지 않고 문서 안에서 서술된 정보가 완결성을 가지도록 작성할 필요가 있다. 두 번째로 방대한 문서에서 필요한 정보를 기계적으로 빠르게 읽고 처리하여(정보 처리의 효율성) 독자가 소화하기 쉬운 형태로 가공할 수 있는 체계를 문서 파일에 심어놓을 필요가 있다. 이를 위해서는 표준 한국어 인코딩(encoding) 체계에 포함되지 않은 특수문자 등의 사용을 줄여 기계 가독성(machine readability)을 높이고, 수치 정보를 문장에서 서술하기보다 표로 정리하는 비중을 높이는 것을 고려할 필요가 있다.

스크립트 언어에 대한 적응도를 높인 공시자료는 다시 말하면 개별 항목들이 찾기 쉽도록 체계적으로 분류되어 있고 항목 사이의 논리적 연계성이 뚜렷한 문서이다. 이는 인간의 관점에서도 그렇지 않은 문서와 비교하였을 때 상대적으로 빠르게 읽고 이해할 수 있을 것이다. 실제로 Allee et al.(2018)의 연구에 따르면 스크립터 적응성이 높은 공시자료를 배포한 기업일수록 새로운 정보의 확산 속도가 빨라 가격이 효율적으로 조정되는 경향을 보였다. 더불어 XBRL의 도입으로 재무 공시의 스크립트 적응성이 높아짐에 따라 영업이익 발표 이후 주가의 지연반응(post-earnings announcement drift)이 줄어들고(Efendi et al., 2014) 재무분석가의 예측 정확도가 향상하여(Felo et al., 2018) 기계 가독성이 높은 문서에서 인간 독자들도 빠르게 정보를 습득할 수 있다는 점을 시사하고 있다.

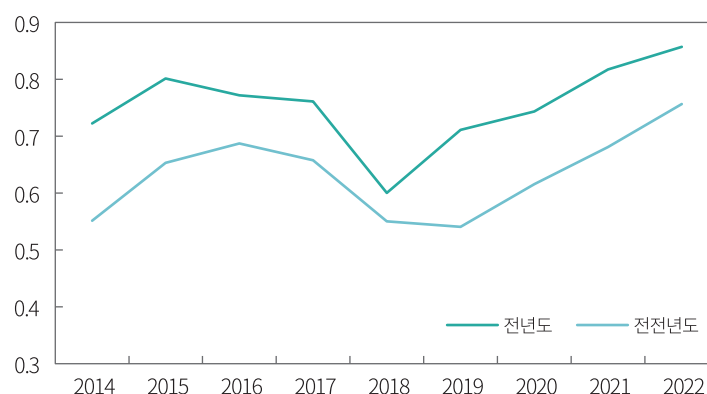
정보 기반 투자 활성화에 대한 시사점

재무 공시의 존재 이유가 투자자에게 포괄적이고 검증된 정보를 전달하는 것인 만큼 구조화 공시와 이에 대한 디지털 접근성을 높이는 것은 정보의 부재 또는 비대칭을 해소하고 가치에 기반한 투자를 활성화하는 것에 궁극적인 목표를 두어야 한다. 이러한 관점에서 AI의 시대에 디지털 구조화 공시제도의 발전 방향에 대한 시사점을 크게 세 가지로 제시해 볼 수 있다.

첫 번째는 온라인 접근성의 제고가 반드시 기계 가독성의 향상을 의미하지는 않는다는 사실이다. 앞서 설명하였듯이 가독성 높은 공시자료란 내용의 논리적 구조가 명확하여 정보의 탐색과 처리가 쉽도록 다층적으로 구획된 문서를 의미한다. 공시자료에 쉽게 접근할 수 있도록 온라인 아카이브에 저장하고 API를 통해 배포하는 등의 인프라는 효과적으로 구조화된 공시자료에 대한 접근성을 높일 뿐, 공시자료의 가독성을 높이는 문제와는 별개의 과제로 보아야 할 것이다.

두 번째로 기계 가독성이 높은 공시 문서가 반드시 정보의 효율적인 확산을 보장하지는 않는다는 점이다. 국내 상장기업의 사업보고서 중 (연결)재무제표 및 주석에 포함된 정보는 지난 회계연도와 비교하였을 때 유사성⁴⁾이 매우 높은 것으로 보인다(〈그림 5〉 참조). 일반적이고 중복되는 정보(boilerplate)가 많아 전년도 공시와 내용이 비슷할수록 투자자는 이를 새로운 정보로 받아들이지 않고 따라서 시장가격의 반응성 또한 낮은데(Brown & Tucker, 2011), 이는 투자자가 재무 공시에서 새롭거나 유용한 정보를 추출하는 비용이 지나치게 높아지기 때문이다(Blankespoor et al., 2020). 따라서 불필요하게 반복되는 정보가 많은 재무 공시는 정보의 효율적인 확산이라는 디지털 구조화 공시의 지향점에 반한다고 볼 수 있다.

〈그림 5〉 사업보고서의 연도별 유사도



주 : 연결재무제표 및 주석에 포함된 단어를 분석하여 직전 1~2 회계연도 대비 자카드 유사도(Jaccard Similarity)를 측정하였으며 90% 이상의 문서에서 사용된 단어(예를 들어 '총자산')는 배제하였다.

자료: 금융감독원 전자공시시스템 및 저자 계산

4) 두 문서의 유사성을 측정하는 지표로 자카드 유사도(Jaccard Similarity)를 사용하였다. 예를 들어 올해 사업보고서에 포함된 단어의 집합을 A, 전년도를 B라고 했을 때, 자카드 유사도는 A와 B의 교집합에 포함된 단어의 수를 합집합에 포함된 단어의 수로 나눈 비율로 정의된다. 두 보고서가 완전히 같은 경우 자카드 유사도는 1, 완전히 다른 경우 0의 값을 가지며 공통으로 사용된 단어의 비중이 높을수록 1에 가까운 값을 가지게 된다.

마지막으로 공시자료의 정보성과 접근성 개선에 투자되는 비용은 초기에 발생하나 그 효용은 장기에 걸쳐서 실현된다는 점이다. 미국의 사례에서 볼 수 있듯이 XBRL 도입 초기 기업은 규정을 준수하는데 큰 비용을 부담할 수 있으며 이는 특히 관련된 경험과 기술적 지원이 부족한 중소기업에게 가증되는 경향을 보인다(SEC, 2023). 더불어 XBRL 도입 초기에는 구조화된 정보를 효율적으로 처리할 수 있는 자원이 풍부한 대규모 기관 투자자를 중심으로 비대칭적인 정보의 확산이 발생할 수 있다(Blankespoor et al., 2014). 따라서 단기적인 비용과 수혜의 불일치 및 비대칭성을 해소하기 위하여 투자자, 기업, 정부 사이의 지속적인 소통과 정책적 지원으로 공시자료의 활용도를 높일 필요가 있을 것이다.

참고문헌

- 금융감독원, 2023. 3. 31, 사업보고서 등을 통해 공시되는 상장사 및 비상장사의 재무정보가 국제표준(XBRL) 데이터 기반으로 전면 개편됩니다, 보도자료.
- 홍지연, 2023, 국내 기업의 XBRL 재무공시 의무화 확대 의의와 과제, 자본시장연구원 『자본시장포커스』 2023-10.
- Allee, K. D., DeAngelis, M. D., Moon Jr, J. R., 2018, Disclosure “scriptability”, *Journal of Accounting Research* 56(2), 363-430.
- Blankespoor, E., deHaan, E., Marinovic, I., 2020, Disclosure processing costs, investors’ information choice, and equity market outcomes: A review, *Journal of Accounting and Economics* 70(2-3), 101344.
- Blankespoor, E., Miller, B.P., White, H.D., 2014, Initial evidence on the market impact of the XBRL mandate, *Review of Accounting Studies* 19, 1468-1503.
- Brown, S.V., Tucker, J.W., 2011, Large-sample evidence on firms’ year-over-year MD&A modifications, *Journal of Accounting Research* 49(2), 309-346.
- Cao, S., Jiang, W., Yang, B., Zhang, A. L., 2023, How to talk when a machine is listening: Corporate disclosure in the age of AI, *The Review of Financial Studies* 36(9), 3603-3642.

- Efendi, J., Park, J.D., Smith, L.M., 2014, Do XBRL filings enhance informational efficiency? Early evidence from post-earnings announcement drift, *Journal of Business Research* 67(6), 1099-1105.
- Felo, A.J., Kim, J.W., Lim, J.H., 2018, Can XBRL detailed tagging of footnotes improve financial analysts' information environment? *International Journal of Accounting Information Systems* 28, 45-58.
- Loughran, T., McDonald, B., 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *The Journal of finance* 66(1), 35-65.
- U.S. Securities and Exchange Commission, 2023. 12, Semi-annual Report to Congress Regarding Public and Internal Use of Machine-Readable Data for Corporate Disclosures.
- Wiggleworth, R., 2020. 12. 5, Robo-surveillance shifts tone of CEO earnings calls, *Financial Times*.

금융감독원 전자공시시스템(DART)	https://dart.fss.or.kr
EDINET	https://disclosure2.edinet-fsa.go.jp/
SEC EDGAR	https://www.sec.gov/edgar